



**DIRECTIONS IN DEVELOPMENT**  
**Human Development**

# Analyzing Markets for Health Workers

*Insights from Labor and  
Health Economics*

Barbara McPake, Anthony Scott, and Ijeoma Edeka



THE WORLD BANK



## Analyzing Markets for Health Workers



**DIRECTIONS IN DEVELOPMENT**

Human Development

# Analyzing Markets for Health Workers

## *Insights from Labor and Health Economics*

Barbara McPake, Anthony Scott, and Ijeoma Edoka



**THE WORLD BANK**  
Washington, D.C.

© 2014 International Bank for Reconstruction and Development / The World Bank  
1818 H Street NW, Washington, DC 20433  
Telephone: 202-473-1000; Internet: [www.worldbank.org](http://www.worldbank.org)

Some rights reserved

1 2 3 4 17 16 15 14

This work is a product of the staff of The World Bank with external contributions. The findings, interpretations, and conclusions expressed in this work do not necessarily reflect the views of The World Bank, its Board of Executive Directors, or the governments they represent. The World Bank does not guarantee the accuracy of the data included in this work. The boundaries, colors, denominations, and other information shown on any map in this work do not imply any judgment on the part of The World Bank concerning the legal status of any territory or the endorsement or acceptance of such boundaries.

Nothing herein shall constitute or be considered to be a limitation upon or waiver of the privileges and immunities of The World Bank, all of which are specifically reserved.

### ***Rights and Permissions***



This work is available under the Creative Commons Attribution 3.0 Unported license (CC BY 3.0) <http://creativecommons.org/licenses/by/3.0>. Under the Creative Commons Attribution license, you are free to copy, distribute, transmit, and adapt this work, including for commercial purposes, under the following conditions:

**Attribution**—Please cite the work as follows: McPake, Barbara, Anthony Scott, and Ijeoma Edoka. 2014. *Analyzing Markets for Health Workers: Insights from Labor and Health Economics*. Directions in Development. Washington, DC: World Bank. doi: 10.1596/978-1-4648-0224-9. License: Creative Commons Attribution CC 3.0 IGO

**Translations**—If you create a translation of this work, please add the following disclaimer along with the attribution: *This translation was not created by The World Bank and should not be considered an official World Bank translation. The World Bank shall not be liable for any content or error in this translation.*

**Adaptations**—If you create an adaptation of this work, please add the following disclaimer along with the attribution: *This is an adaptation of an original work by The World Bank. Responsibility for the views and opinions expressed in the adaptation rests solely with the author or authors of the adaptation and are not endorsed by The World Bank.*

**Third-party content**—The World Bank does not necessarily own each component of the content contained within the work. The World Bank therefore does not warrant that the use of any third-party-owned individual component or part contained in the work will not infringe on the rights of those third parties. The risk of claims resulting from such infringement rests solely with you. If you wish to re-use a component of the work, it is your responsibility to determine whether permission is needed for that re-use and to obtain permission from the copyright owner. Examples of components can include, but are not limited to, tables, figures, or images.

All queries on rights and licenses should be addressed to the Publishing and Knowledge Division, The World Bank, 1818 H Street NW, Washington, DC 20433, USA; fax: 202-522-2625; e-mail: [pubrights@worldbank.org](mailto:pubrights@worldbank.org).

ISBN (paper): 978-1-4648-0224-9

ISBN (electronic): 978-1-4648-0225-6

DOI: 10.1596/978-1-4648-0224-9

*Cover photo:* © Chhor Sokunthea / World Bank. Used with permission. Further permission required for reuse.

### **Library of Congress Cataloging-in-Publication Data**

McPake, Barbara, author.

Analyzing markets for health workers : insights from labor and health economics / Barbara McPake, Anthony Scott, and Ijeoma Edoka.

p. ; cm. — (Directions in development)

Includes bibliographical references.

ISBN 978-1-4648-0224-9 (paper: alk. paper) — ISBN 978-1-4648-0225-6

I. Scott, Anthony, 1967- author. II. Edoka, Ijeoma, author. III. World Bank, issuing body. IV. Title.

V. Series: Directions in development (Washington, D.C.)

[DNLM: 1. Health Manpower—economics. 2. Health Personnel—economics. W 76.1]

RA410.7

331.7'6161069—dc23

2014008441

# Contents

<i>Preface</i>		<i>vii</i>
<i>Acknowledgments</i>		<i>ix</i>
<i>About the Authors</i>		<i>xi</i>
<i>Abbreviations</i>		<i>xiii</i>
	<b>Overview</b>	<b>1</b>
<b>Chapter 1</b>	<b>Introduction</b>	<b>3</b>
	Structure of the Publication	6
	References	7
<b>Chapter 2</b>	<b>A Framework for Analysis</b>	<b>9</b>
	The Role of the Health Workforce in the Health Care System	9
	Market Forces and Market Failure in Health Workers' Labor Markets	12
	References	16
<b>Chapter 3</b>	<b>Policy Interventions and the Health Workforce: Economic Issues and Evidence</b>	<b>17</b>
	Demand for Health Workers	17
	Supply and Productivity of the Health Workforce	27
	Note	48
	References	48
<b>Chapter 4</b>	<b>Analytical Approaches</b>	<b>59</b>
	Descriptive Labor Market Analysis: The Nature and Extent of Disequilibrium	60
	Causal Labor Market Analysis	65
	Data Requirements	70
	References	73
<b>Chapter 5</b>	<b>Gaps and Future Research</b>	<b>75</b>
	Gaps by Topic	75

Geographic Gaps	78
The Way Forward	78
Note	79
References	79

## Boxes

1.1	Labor and Health Economics	4
2.1	Labor Productivity	11
3.1	Pay Flexibility—The Theory of Compensating Differences	19
4.1	MABEL Panel Survey of Doctors	72

## Figures

1.1	Count of Studies in Countries Covered by the Review	7
2.1	The Role of the Health Workforce in the Health Care System	10
B4.1.1	Types of Data Collected by the MABEL Survey	72



## Preface

Improving equitable access to quality health services is one of the main pillars of the World Bank Health, Nutrition, and Population Strategy. The Bank gives high priority to ensuring equitable and sustainable improvements in health outcomes with particular attention to enhancing the well-being of the poor and vulnerable population as part of its primary mission to reduce poverty and promote shared prosperity across the globe. Within this framework, the Bank supports the aspirations of developing countries toward universal health coverage as an important goal that will contribute to each country's efforts in ensuring inclusive and sustainable development.

The Bank has identified the inadequate availability of health services and health workers, especially in rural and remote areas, as well as weak management and limited incentives—often not linked to performance—as some of the leading causes of the poor performance of health systems. The Human Resources for Health (HRH) program at the World Bank has been established to assist countries to carry out critical upstream analytic work that will inform health policy and improve the performance of health systems in an equitable and sustainable manner. The focus of the HRH program is on areas where the World Bank has a comparative advantage, including labor market analysis, the synergies between HRH and health financing policies, HRH budget and cost analysis, and assessment of health worker incentives and evaluation of performance-based pay policies.

This publication is part of the Bank's multiyear program to enhance its knowledge of HRH policies. The program's ultimate objective is to strengthen knowledge and capacity to collect evidence, analyze, and evaluate the effectiveness of HRH interventions in the context of a country's health system strengthening strategy. It specifically addresses the theoretical and empirical evidence on health labor markets in low- and middle-income countries.

Health labor market analysis has much to contribute to resolving globally widespread HRH problems, and continuing neglect of these problems provides some explanation for their persistence. Policy makers in countries promulgating or refining strategies for achieving universal health coverage will find it important to understand how key elements in their health labor market are likely to interact and how these interactions could help—or hinder—progress toward universal health coverage. These interactions are complex and multidimensional, and this publication highlights some areas where forces in the health labor market matter most.



## Acknowledgments

The authors would like to thank the following persons for their contributions to the publication: Edson Araújo (task team leader, Health, Nutrition, and Population [HNP] Global Practice, World Bank Group), Akiko Maeda (HNP Global Practice, World Bank Group), Christophe Lemiere (HNP Global Practice, World Bank Group), Giorgio Commeto (Global Health Workforce Alliance), and Atef El Maghraby (African Development Bank), Daniela Hoshino (HNP Global Practice, World Bank Group), and Jonathan Aspin (editor, World Bank consultant). The publication greatly benefited from comments received from Christopher Herbst (HNP Global Practice, World Bank Group), Michael Weber (Social Protection and Labor Global Practice, World Bank Group), and Wanda Jaskiewicz (Health Workforce Performance, CapacityPlus). The authors also thank participants at the Regional Workshop on Health Labor Market Analysis, Tunisia, March 2013, especially Dr. Atef El Maghraby and his team at the African Development Bank.

This publication was supported by the generous grant provided by the Norwegian Ministry of Foreign Affairs and the Norwegian Agency for Development Cooperation (NORAD) through the Multi-Donor Trust Fund Strengthening Human Resources for Health Policy in Developing Countries.



## About the Authors

**Barbara McPake**, BA, PhD, is a health economist specializing in health policy and health systems research. She has 24 years of experience in these areas based in three U.K. university departments and is currently director of the Institute for International Health and Development at Queen Margaret University, Edinburgh, Scotland. Her research interests have included work on financing policy, using contracts and “pay for performance,” understanding the implications of hospital reforms, and looking at the informal economies of the health workforce. She was program director of the Health Systems Development Knowledge Programme (2001–06), and is currently one of two research directors of the ReBuild Research Programme Consortium (2011–17), focusing on reestablishment of health systems in postconflict settings (both funded by the U.K.’s Department for International Development [DFID]).

**Anthony Scott**, PhD, leads the Health Economics Research Program at the Melbourne Institute of Applied Economic and Social Research at the University of Melbourne, and jointly coordinates the University of Melbourne Health Economics Group. He has a PhD in economics from the University of Aberdeen. Dr. Scott is a principal research fellow at the National Health and Medical Research Council (NHMRC) and associate editor of the *Journal of Health Economics*. He leads the Centre of Research Excellence in Medical Workforce Dynamics ([www.mabel.org.au](http://www.mabel.org.au)). Funded by the NHMRC, the Centre runs Medicine in Australia: Balancing Employment and Life (MABEL), a large, nationally representative panel survey of physicians. Dr. Scott’s research interests focus on the behavior of physicians, the health workforce, incentives and performance, and primary care.

**Ijeoma Edoka**, PhD, is a health economist and is currently a research fellow at the Institute for International Health and Development, Queen Margaret University, Edinburgh. Her research interests lie in empirical analysis of health policies in low- and middle-income countries. She is currently involved in a number of projects evaluating the impact of health financing policies in Cambodia, Sierra Leone, Uganda, and Zimbabwe using advanced econometric methods. She has also led a team of health economists on the economic evaluation of community health worker programs in Ethiopia, Indonesia, and Kenya.



## Abbreviations

BRIC	Brazil, the Russian Federation, India, and China
DCEs	discrete choice experiments
EQUINET	Regional Network for Equity in Health in East and Southern Africa
FTE	full-time equivalent
IT	information technology
IV	instrumental variable
LMICs	low- and middle-income countries
MABEL	Medicine in Australia: Balancing Employment and Life
NHS	National Health Service
OECD	Organisation for Economic Co-operation and Development
WHO	World Health Organization





# Overview

The purpose of this publication is to provide an overview of the key issues when attempting to apply economics to the analysis of health workers' labor markets. Though much has been written and planned about health human resources, a major weakness with most of this analysis is that it does not use an economic perspective. The use of an explicit economic framework applied by trained economists moves the focus away from simplistic but costly policy responses such as training more doctors and nurses, toward understanding more carefully the role of incentives, productivity, and the distribution of health workers. The health workforce is but one part of the health system, and a focus of analysis on only the health workforce is insufficient to be able to determine the optimal number of health workers. Market forces cannot be relied upon to solve health worker shortages or maldistribution, due to well-recognized market failures in health care. This also has implications for how labor economics and labor market analysis can be applied and used successfully in the health care sector.

The policy drivers of health workforce reform are a seemingly persistent mismatch between “need” and existing supply of health care services, including health workers. The document outlines how the health workforce is related to the rest of the health care system, and to other social and economic determinants of population health and well-being. It addresses the question of why market forces cannot be relied on to solve health worker shortages and why government intervention and regulation are required due to broader and well-recognized market failures in health care and health labor markets. The scale and types of government intervention vary across countries. The publication summarizes the evidence from low- and middle-income countries of market failures, government interventions, and their implications for health care provision.

The document combines the analytical framework of labor economics with an understanding of market failure provided by health economics, to provide a framework that can be used to further understanding of the dynamics of health worker labor markets. Demand-side issues include pay-setting arrangements and skill mix and task substitution. Supply-side issues include education and training, workforce participation, migration and retention, dual practice, geographic

distribution, productivity and performance, and health worker motivation. The economic approach to each of these issues is summarized, followed by a brief review of the literature in high income countries, and a more detailed review of studies from low- and middle-income countries.

Applying the economic framework to policy issues requires an understanding of both the types of economic analysis that can be conducted, and also the data required to undertake such analyses. There are two main types of economic analysis that can be conducted. Descriptive labor market analysis examines the current state and trends in the labor market and might generate more specific research questions and hypotheses. Causal labor market analysis is concerned with examining the causal effect of factors influencing the labor market behavior of employers and health workers, and can be based on the evaluation of policy changes on a range of labor market outcomes.

To conduct these types of analyses, a range of different sources of data can be used. However, data must include the earnings of health workers, and also be a panel/longitudinal, such that health workers data can be linked over time. In addition to randomized trials, panel data provides the most powerful data that be used to examine the causal effects of policy and of factors influencing behavior. These are the two most important gaps in current data collections that need to be addressed and that are currently preventing the application of economics to the analysis of health worker labor markets. Better data will also attract those with economics and micro-econometrics training to conduct research in this area.

There are three essential ways forward: first, more systematic and consistent application of economic thinking to human resources for health issues; second, building capacity by investment in better administrative and survey data that are matched to each other, and include data on earnings; and third, building capacity by involving more health and labor economists in health workforce research.

# Introduction

*The aim of this publication is to examine how labor and health economics can be used to analyze and better understand the role and functions of health worker labor markets. Health workforce shortages stem not only from inadequate overall supply, but also from suboptimal allocation of health human resources by location and role. Low performance and productivity are also issues. These three problems are often compounded by a resource problem—the gap between the finances required for an “adequate” workforce and those likely available. The application of labor economics to health care labor markets needs to account for the specific institutional features and market failures in health care.*

Policy responses to shortages of health workers in low- and middle-income countries (LMICs) have to date almost exclusively focused on addressing shortages through “scaling up” interventions that increase the supply of health workers. This assumes that more health workers are a cost-effective way to improve the population’s health.

Though training and numbers are clearly an issue, it is also vital to ensure that the health workers already employed are used to their best effect, are productive, and are employed at reasonable cost; and that those newly trained are retained and encouraged to provide cost-effective treatments and procedures in specialties and geographic areas where the need for health care is high. These should be key objectives of health human resource policy, taking account of the ethical and equity issues surrounding health workforce migration between countries and between urban and rural areas within a country.

Pure scaling up largely ignores the potential contribution of labor and health economics in understanding how health worker labor markets function. An economic approach to labor markets is fundamental in fully understanding issues of health workforce shortages, productivity, and performance, and the appropriate policy responses.

The issues can be categorized into four “problems”: quantity, allocation, performance, and resources (Andalon and Fields, forthcoming). Health workforce shortages are due not only to inadequate overall supply (the quantity problem),

but also to a suboptimal allocation of health human resources in a range of interdependent submarkets (the allocation problem), notably between:

- geographic areas within and between countries, including urban–rural imbalance and migration;
- public and private sectors, including issues arising from dual practice;
- medicine, nursing, and other health workers—skill mix;
- medical specialties—generalist, primary, community-based care versus specialist care; and
- treatment settings—primary care, outpatients, acute care hospitals, and informal care at home.

Low performance and productivity are also often issues. If they were increased, fewer health workers would be required and health outcomes improved.

These three issues are often compounded by the resource problem—the gap between the finances required to expand the workforce to the required degree and those available in the near future.

The aim of this document is to examine how labor and health economics can be used to analyze and better understand the role and functions of health worker labor markets. It draws on the framework of labor economics (Andalon and Fields, forthcoming; Scheffler *et al.* 2012) and the insights of health economics to provide a conceptual framework that can contribute to guiding more appropriate and effective analysis and data collection related to the health workforce. The conceptual framework of labor economics has been highlighted in publications by the World Bank and World Health Organization (WHO) (Andalon and Fields, forthcoming; Scheffler *et al.* 2012). This publication goes several steps further by reviewing the types of analysis and data requirements necessary to apply this framework to health workforce issues in LMICs in more depth than has been undertaken previously.

The application of labor economics to the analysis of general labor markets is often different from the application of labor and health economics to the analysis of health workers' labor markets (box 1.1). This is largely due to the different set of policy and institutional issues that drive theoretical and empirical research

---

### **Box 1.1 Labor and Health Economics**

Labor economics is a large field of economics that provides a framework for understanding how labor markets work. Only a handful of its concepts and tools, however, have been applied to health worker labor markets.

The theoretical and empirical approaches in traditional labor economics deal with issues and market imperfections in aggregate, that is, the whole labor market. Some of these topics are less relevant to health workers as they focus on low-income workers and unemployment, minimum wages, wage inequality, and trade unions.

---

*box continues next page*

**Box 1.1 Labor and Health Economics** *(continued)*

Other topics have been little analyzed in health care because the issues faced by health workers are assumed to be no different from those faced by other workers. Examining one or two specific health care occupations is unlikely to inform broad economic policy. These topics include gender discrimination, immigration, human capital development, and skills formation. Such topics are largely driven by broader economic objectives, notably raising employment and productivity, and thus economic growth.

The drivers of, and therefore interest in, health workforce policy are different. Though the benefits of achieving these broader economic objectives are also important for a productive health care sector, improvements in health status are valued independently of these objectives. Better health status indeed leads to gains in employment and economic growth, but is valued in its own right as a separate objective, sometimes called “extra-welfarism” in the health economics literature (Hurley 2000).

The policy drivers of health workforce reform in health care are a seemingly persistent mismatch between “need” and existing supply of health care services, including health workers. The issues are broader than a focus on health worker labor markets and include the efficiency and distribution of health care organizations. This leads to a natural focus on health worker labor markets given the labor intensive nature of health care production. As we will see in chapter 3, because demand cannot easily be defined in health care due to market failure, concepts of medically defined need dominate, which may be unrelated to willingness to pay and patients’ preferences. This leads to the impression in health care that there are never enough resources to meet this need, and that all needs should be met. There is little consideration of resource scarcity.

These particular features of labor and product markets in health care have meant that the analysis of health labor markets has taken a different direction from that of wider labor markets. For example, asymmetry of information between doctors and patients has led, in health economics, to a focus on principal–agent relationships, optimal incentive contracts, specific areas of information and organizational economics, and the effect of different methods of remuneration on productivity and performance, rather than the level of wages (a central focus of labor economics).

Asymmetry of information has also led to a focus on doctors. As doctors are self-employed in many countries, traditional labor market analyses of salaried workers and labor demand and supply have been supplanted by the theoretical and empirical analysis of the behavior and productivity of groups of health workers in teams and firms, in, for example, private medical groups (small firms) or hospitals. A key practical aspect of labor market analysis in a particular country is therefore whether workers are salaried (usually in the public sector) or self-employed (in the private sector).

---

in labor and health economics. Often the application of labor economics to health care labor markets ignores the particular institutional and market features of health care labor markets. What labor and health economics do share is a common set of microeconomic theory and micro-econometric empirical methods and tools to analyze labor market issues.

## Structure of the Publication

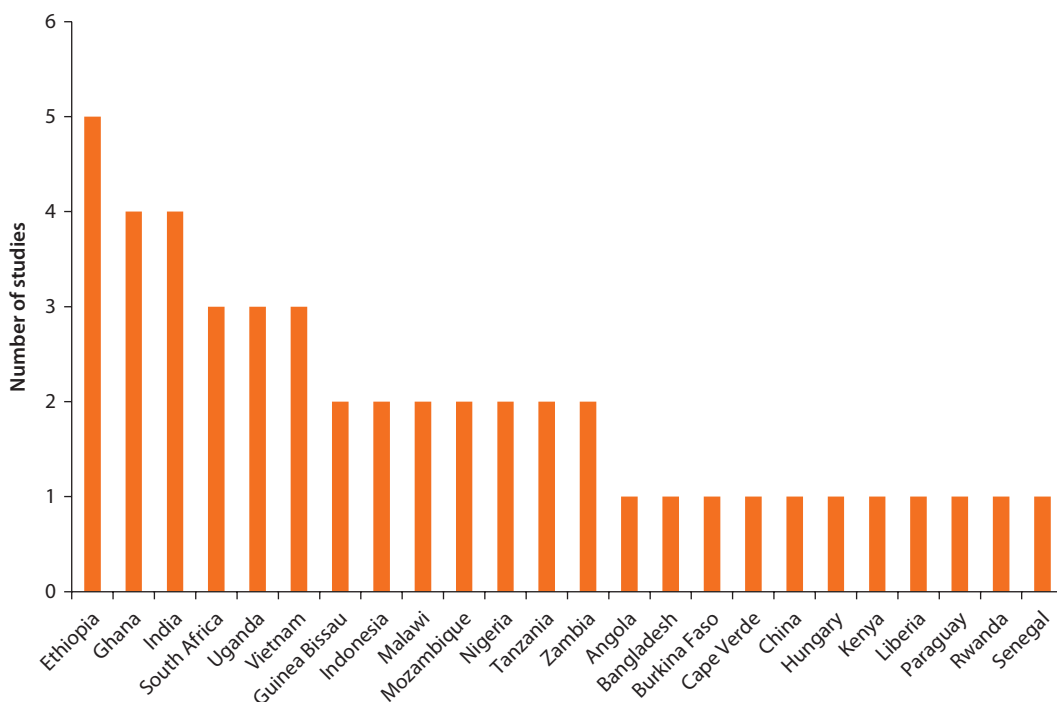
The document has four main chapters. After this introduction, the second section sets out a broad framework that needs to be used when examining health care labor markets. This includes outlining how health workers are related to the rest of the health care system, and to other social and economic determinants of population health and well-being. The purpose here is to show that the health workforce is only one part of improving population health and well-being, and that there may be other more cost-effective ways to improve health than by scaling up the health workforce. We then address the question of why market forces cannot be relied on to solve health worker shortages. Government intervention and regulation are required due to broader and well-recognized market failures in health care and health workforce labor markets, though the scale and types of government intervention can vary across countries for a range of reasons.

The third section summarizes the key issues surrounding the demand and supply of health workers and how these interact in the health worker labor market. Demand-side issues include pay-setting arrangements and skill mix and task substitution. Supply-side issues include retention, geographic distribution, sectoral distribution (including dual practice), and performance and motivation. Discussion of these issues is followed by a brief summary of the evidence from health care labor markets in high-income countries, and a more detailed summary of the evidence from LMICs.

The review of evidence from LMICs was undertaken using a search strategy to identify recent work in English that could broadly be defined as using health labor market analysis or elements of it in LMIC contexts. A primary focus is the recent literature so as to reflect an interest in the current depiction of health labor markets, recognizing that these, and the factors influencing them, change quite rapidly. Databases searched included Science Direct, Google Scholar, World Bank archives, the Health Systems 2020, and HRH Global Resource Center databases. The date range searched was 2006 to 2013, although some references prior to 2006 were identified through the citations of those articles as particularly relevant.

A count of the geographic focus of studies from LMICs used in this document (and listed in the References) is in figure 1.1. This classification is quite rough: different elements of some studies are reported in more than one paper, and papers that covered a general class of countries (for example “developing”) or a whole continent were excluded from the count, while ones covering a distinctive list of countries or areas within one or more countries were included. Our focus is primarily on LMICs and it is not surprising that the bulk of this literature is focused on countries that are deemed to be in human resource crisis, usually according to the standard advanced by *The World Health Report 2006*, of which the great majority are in Africa (WHO 2006). It is also unsurprising that in a review restricted to English language publication, Anglophone countries are more strongly represented than Francophone and Lusophone ones.

The fourth section summarizes the broad analytical approaches used in economics, focusing on issues of causality and labor market dynamics. This also

**Figure 1.1 Count of Studies in Countries Covered by the Review**

**Note:** 47 studies with geographically specific empirical content relating to low- and middle-income countries were identified (see References). These are spread across 23 countries—17 in Africa.

helps to define the data requirements that are essential to be able both to describe the current state of the health worker labor market and to evaluate the causal impact of policies that attempt to restore equilibrium.

The final section suggests some gaps in research and analysis for health worker labor markets in LMICs.

## References

- Andalon, M., and G. Fields. Forthcoming. "A Labour Market Approach." In *Human Resources for Health in Africa: A New Look at the Crisis*, edited by R. Scheffler and A. Soucat. Washington, DC: World Bank.
- Hurley, J. 2000. "An Overview of the Normative Economics of the Health Sector." In *Handbook of Health Economics*, edited by J. C. Anthony and P. N. Joseph, 55–118. Elsevier.
- Scheffler, R., T. Bruckner, and J. Spetz. 2012. *The Labour Market for Human Resources for Health in Low and Middle Income Countries*. Human Resources for Health Observer, Issue No. 11. Geneva: World Health Organization.
- WHO (World Health Organization). 2006. *The World Health Report: Working Together for Health*. Geneva: WHO.





# A Framework for Analysis

*Health workforce policy is a means to an end, and making decisions about the health workforce in isolation from the rest of the health care system is problematic.*

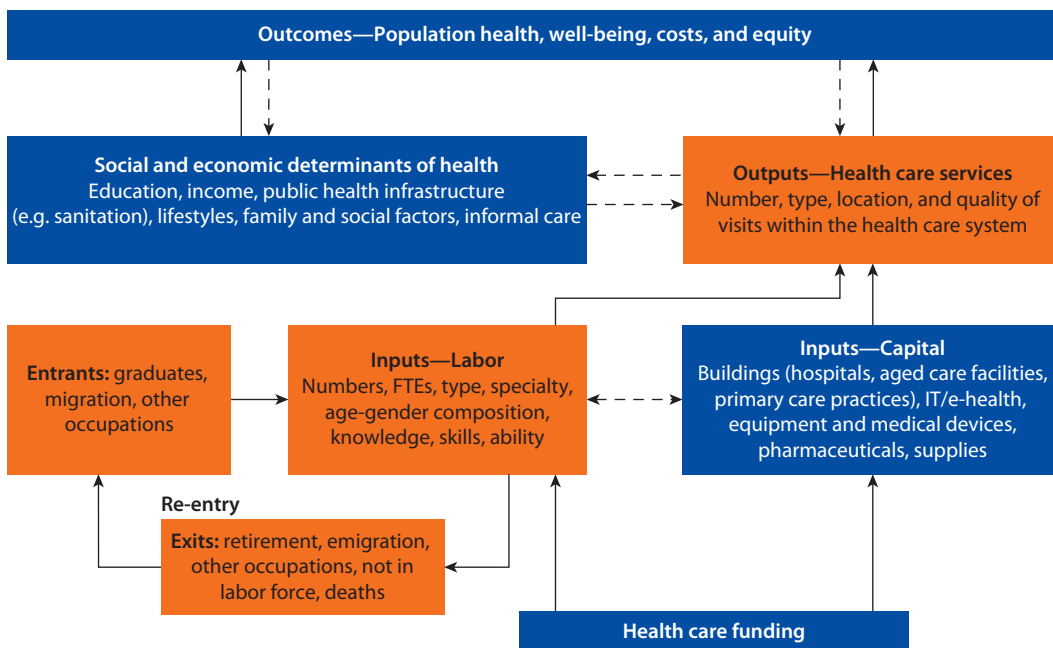
*The role of the health workforce in the health care system can be illustrated using a “production function” framework. The framework recognizes that the ultimate aims of interventions in health worker labor markets are to improve population health with limited resources and to achieve equity goals.*

*Using a production function framework suggests that size, distribution, and productivity of the health workforce can only be properly examined by accounting for all interrelated factors. But the ability to model these relationships empirically using good data sources requires further developments in methods and data sources for labor market analysis, as modeling the whole system is highly complex—thus modeling parts is often pragmatic.*

## The Role of the Health Workforce in the Health Care System

Decisions about the health workforce should not be made in isolation from the rest of the health care system, given the tight linkages shown all the more clearly in a production function framework (figure 2.1). This framework specifies inputs, outputs, and outcomes (including costs). It recognizes that the ultimate aims of interventions in health worker labor markets are to improve population health with limited resources and to achieve equity goals, in a trade-off between efficiency and equity.

The key issue is that the size, distribution, and productivity of the health workforce can only be properly examined by accounting for all factors in the framework. The ability to model these relationships empirically using good data sources helps to define how the methods and data sources of labor market analysis should be developed. However, modeling the whole system is extremely complex and so modeling parts of it is the pragmatic solution. When using a partial approach, it is important to recognize that modeling only a limited part of the system inevitably leads to assumptions being made about the rest of the system, and these need to be explicit.

**Figure 2.1 The Role of the Health Workforce in the Health Care System**

Source: Scott *et al.* 2011.

Note: FTE = full-time equivalent; IT = information technology.

The key assumption is that all other aspects of the system do not change. For example, many approaches to health workforce planning, including supply and demand projections, justify scaling up yet focus only on the orange parts of the figure (Scott *et al.* 2011). They ignore the effects of increasing health workforce supply on health care costs and population health. Thus decisions are being made without information on the costs or benefits of these large investments, or which parts of the population receive the benefits and bear the costs. These issues are rarely explicitly addressed.

The framework in figure 2.1 recognizes that the population's health and well-being are influenced by a variety of interrelated and dynamic factors, of which the health workforce is but one (Dussault *et al.* 2010). The first point raised by the figure is that fewer health workers may be required if health is best improved through improved sanitation and other public health infrastructure, or through basic education about disease prevention (see the *Social and economic determinants of health* cell), i.e. many potential policies to improve health do not directly involve the health care system. A focus on basic public health measures may reduce the need for health workers to treat disease in the future.

The second point is that the relationship between the number of health workers and population health is not straightforward. First, there will be diminishing marginal returns for each additional dollar spent on health workers. One might expect the marginal benefit (in health outcomes, for example) from training

an additional doctor in the United States to be lower (and the marginal cost much higher) than the same investment in training an additional doctor in Malawi, assuming the doctor does not migrate from in Malawi. Second, there is an assumption that all treatments and procedures supplied by additional health workers are effective. In fact, a large proportion of health care services have no evidence base for their effects on health or on costs (Berwick and Hackbarth 2012). More health professionals undertaking the same potentially ineffective interventions may not lead to improvements in population health but may still substantially increase health care costs.

At the bottom of the figure, health care funding—including the various sources of funding and payment mechanisms—influences the quantity of inputs and their productivity. How the system is organized defines how inputs are combined within health care organizations, such as hospitals and primary care settings. The demand for health workers is determined by these organizations, which in turn are influenced by the needs of patients and the role of health care financing in translating need into effective demand.

Health workers require inputs such as physical infrastructure, equipment, support, and training to provide effective care. These inputs need to be combined with other inputs and embedded in organizational structures. Each input has its own market that determines the number of inputs and their productivity, of which labor is the most important (box 2.1).

---

### **Box 2.1 Labor Productivity**

Labor productivity refers to the productivity of health workers and is expressed as the ratio of health care outputs (e.g. number of visits) to labor inputs (e.g. hours worked, full-time equivalents [FTEs], or number of workers).

An important concept is the marginal product of labor, defined as the effect of a unit change in inputs (e.g. an extra hour worked, extra FTE, or extra worker) on output (e.g. the number of visits). An additional hour worked, for example, may lead to the health professional seeing three additional patients. The marginal value product is the dollar value of this additional output, i.e. what employers and patients are willing to pay for the extra output. In a well-functioning market, the value of output defines the wage rate of health workers and the prices of other inputs.

In health care, the concept of the marginal value product is more difficult to define and measure since outputs are not valued in monetary terms, but are nevertheless valued by patients as they influence quality of care and health outcomes. Market failures mean that the link between the value of services provided to patients and wages is, at best, weak (see following subsection).

Labor productivity (and the marginal product of labor) is not only influenced by hours worked but also by a range of other factors, including the age and gender of workers, their education, skills, abilities, and the level and method of payment. Some health professionals

---

*box continues next page*

**Box 2.1 Labor Productivity** *(continued)*

can see more patients in an hour because they may be more experienced, or because they are paid fee for service rather than a salary.

Labor productivity can also change because of new ways of using and combining existing labor and capital inputs (a different production process) to provide more outputs, or where existing and new labor and capital inputs are combined in such a way as to require fewer inputs to produce a similar level of output. The productivity of a health worker is also influenced by the productivity of other health workers, especially since they usually work in teams. So although the skills and aptitude of a health worker may not alter, his or her productivity may be influenced by new technology or by changing productivity among colleagues.

---

Changes in input mix can also influence the cost of health services without influencing population health and well-being. A different way of combining different inputs may reduce costs while maintaining population health. Such efficiencies can release resources to be used for other more productive purposes. Increases in the costs of inputs (e.g. earnings of the health workforce) mean that fewer other inputs can be used in a system with fixed overall budgets, or may mean that health care costs overall rise with no change in the population's health and well-being.

Some of the relationships in the system are illustrated by the figure's dashed arrows. For example, the number and type of health care services influence the population's health and well-being, but population health and well-being also influence the demand for the number and type of health care services.

Equity of access and issues of financial protection can be examined in the framework by examining the distribution of population health and well-being, health care services, costs, and inputs (hospitals and primary care services) across different population groups (e.g. rural and urban, high and low income).

Figure 2.1, then, provides a descriptive structure of the health workforce and the health care system. But it says little about *how* the labor market for health care services operates—this is addressed in the next subsection.

**Market Forces and Market Failure in Health Workers' Labor Markets**

*Summary: The aim of this subsection is to summarize how labor markets work, and in health care, how market failures can prevent these markets from working well. The building blocks of the market are supply, demand, wages (prices), and market forces that move prices and quantities to equilibrium. In a well-functioning labor market, the demand and supply of labor tend to move to equilibrium, where the market "clears." The wage rate plays a central role because it directly reflects the market value of output produced.*

*Markets can fail because prices are inflexible or because either demand or supply is constrained from responding to price signals. The consequences of market failure mean that government intervention is required to try and ensure good outcomes at reasonable cost. Markets can also fail for reasons related to the nature of demand in health care. The asymmetric information between patients and health workers means defining “demand” is problematic such that wages and payments made to health workers may not reflect consumers’ and employers’ preferences as they would in a well-functioning labor market. This has implications for how labor economics can be applied to health care labor markets.*

*Understanding the operation of these markets therefore helps define which government policies in the health care labor market may be necessary.*

### **Well-Functioning Labor Markets**

In a well-functioning labor market the demand and supply of labor tend to move to equilibrium, where demand equals supply and there are no shortages or surpluses as the market “clears” due to market forces that are driven by changes in the behavior of employers on the demand side and of health workers on the supply side. The building blocks of the market are supply, demand, wages (prices), and market forces that move prices and quantities to equilibrium. In terms of labor demand, employers with a fixed budget can employ fewer workers if wages rise or can employ more workers if wages fall. In terms of labor supply, the number of workers will increase if higher wages are offered. How employers and workers respond to wage signals is therefore a key part of how market forces operate.

For example, if the demand for health workers were to rise, say due to increases in the prevalence of disease, a shortage of health workers would occur. The increase in disease prevalence would be translated into an increased willingness to pay or demand for health workers, as their value in the market rose. This higher willingness to pay would raise the price paid for health workers (their wage), lifting the supply of health workers in two ways.

First, it would provide incentives to increase hours worked for existing health workers and increase workforce participation for health workers currently unemployed or not working for other reasons. Second, it would make the job of health worker more attractive relative to other occupations, leading to an increased supply of those training to become health workers. However, as supply increases and more people decide to become health workers and enter the market, the marginal or additional value of each extra health worker (the wage rate) will fall and the rate at which health workers enter the market will also fall back to the equilibrium wage, where supply once again equals demand.

Market forces therefore operate through the wage rate (price), which reflects the willingness to pay (demand) for a health worker. The wage rate influences how many hours health workers are willing to supply, and how many individuals choose to become health workers relative to entering other occupations.

The wage rate therefore plays a central role in labor economics because it directly reflects the market value of output produced.

### ***Labor Markets in Disequilibrium and Market Failure***

Disequilibrium in well-functioning labor markets described above is short lived. Wage rates that are too high or too low provide incentives to employers and health workers to alter their behavior such that equilibrium is restored. Unemployment (where wages are too high and supply is greater than demand) or shortages (where wages are too low and demand is greater than supply) will be eliminated. But the operation of the market this way, especially of market forces, depends on a number of assumptions which, if they do not hold, can create persistent disequilibriums and inefficiency.

These assumptions are centered on whether wages and the quantity of workers are free to adjust. This in turn depends on the existence of institutions that determine wages; restrictions on entry into the labor market for new workers; workers having the same information about jobs as employers; employers maximizing objectives other than profit; and workers maximizing objectives other than income. Often in these models, a range of other factors are held constant, which is a necessary way to simplify complex behavior, but in empirical work is the main cause of empirical problems.

Markets can fail because they fail to achieve equilibrium, and for reasons related to the nature of demand in health care. Demand for health workers is derived from the demand for health care, in turn derived from the demand for health (see top left of figure 2.1). In the example above of an increase in disease prevalence, it is assumed that those who have the disease know the following: that they have the disease; that health workers can provide an effective treatment likely restore their health and well-being, and so they seek it; and that the price they are willing pay to see a health worker is determined by the value they place on restoring their health.

But patients have much less information than physicians about why they are ill and the value, effectiveness, and quality of the diagnostic options and treatments available (Arrow 1963). This asymmetric information means that observing “demand” is problematic. A principal–agent relationship exists: doctors act as agents for patients and provide diagnosis and have knowledge of the treatment options and their effectiveness. Though doctors can share this information with patients, patients most often place their trust in doctors to make decisions on their behalf, especially when they are ill, have low levels of education, or are elderly and have multiple morbidities. Though higher-income and more educated patients are increasingly informed through the Internet, they still do not possess the skills and experience of medical practitioners.

Thus the demand for health care services and health workers is judged by health workers themselves using judgments of “need” rather than by informed consumers within a constrained budget. No direct link exists between a consumer’s preferences and prices (willingness to pay and market wages), or at best

the link is very weak. Health insurance further weakens price signals. The potential for health workers to recommend care that increases their earnings and reputations but is harmful or ineffective for patients (supplier-induced demand) has led governments to introduce regulation; lengthy periods of training, licensing, and accreditation; and regulations on levels and mechanisms for reimbursing health workers. Such measures also reduce mobility between countries. Ensuring minimum quality standards by restricting entry into domestic and international labor markets also keeps wages relatively high and is another form of market imperfection in health care.

Consequently, wages and reimbursement levels for health workers do not reflect consumers' valuation of the services provided, as they would in a well-functioning market. Since wages (and other prices) are unlikely to reflect the value of output by consumers, health economists use direct measures of quality of care (e.g. mortality rates) and utility-based measures of health outcomes (e.g. quality-adjusted life years) to reflect the value of output. Wages in health care can provide suboptimal market signals to health workers who will supply too many or too few hours, and who will enter or leave the labor market at rates that are not related to changes in demand or even need for health care. When financial incentives (wages) are not linked to consumers' preferences, the decisions made by health workers on specialty choice, rural-urban choices, migration, and treatment decisions will not be optimal. To the extent that governments represent societies' preferences, government intervention is required to address these inefficiencies.

Government steps to curb supplier-induced demand restrict supply, leading to higher wages and inflexibilities in the labor market such that supply takes a long time to respond to changes in demand. Much of this regulation of minimum quality standards is delegated to the health workers' professional organizations that decide on the numbers entering training programs, with the result that supply might be restricted more than optimally. There is therefore a trade-off between high wages and ensuring minimum quality standards for care, though there is little empirical evidence of the effects of licensing (Nicholson and Propper 2011). Subsidizing medical training also contributes to excess demand for medical school places. This excess demand is persistent and does not lead to more places being offered, and arguably maintains physician wages at relatively high levels—another market imperfection.

Some differences between country groups emerge: in low- and middle-income countries, regulation of training and licensing may not be as well established as in high-income countries, and government intervention more broadly may be more limited in less well-developed labor markets, and wages and reimbursement may be more responsive to changes in demand and supply conditions there. This does not mean, however, that these markets work any better or are more efficient and equitable than high-income countries' markets, as information is even more asymmetric, especially when there are few minimum quality standards through licensing.

## References

- Arrow, K. J. 1963. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review* 53: 941–73.
- Berwick, D., and A. Hackbarth. 2012. "Eliminating Waste in US Health Care." *Journal of the American Medical Association* 307: 1513–16.
- Dussault, G., J. Buchan, W. Sermeus, and Z. Padaiga. 2010. "Assessing Future Health Workforce Needs." European Observatory on Health Systems and Policies, World Health Organization, Geneva.
- Nicholson, S., and C. Propper. 2011. "Medical Workforce." In *Handbook of Health Economics*, edited by M. Pauly, T. McGuire, and P. Barros, 873–925. Elsevier.
- Scott, A., P. Sivey, C. Joyce, D. Schofield, and P. Davies. 2011. *Alternative Approaches to Health Workforce Planning*. Report by the National Health Workforce Planning and Research Collaboration, Health Workforce Australia, Adelaide.



# Policy Interventions and the Health Workforce: Economic Issues and Evidence

*The aim of this section is to examine in more detail the demand and supply sides of health worker labor markets, including the empirical evidence of the role and impact of policy interventions—in high-income countries and in more depth in low- and middle-income countries.*

### **Demand for Health Workers**

Demand for health workers in a country is largely determined by what governments and individuals are willing to pay for the health care provided by the health workforce. The concept of willingness to pay is important and distinguishes demand from “need.” Health workforce needs or requirements are often defined by health workforce planners and do not take into account resource constraints, nor the effect of extra health workers on productivity or health outcomes. The demand for the health workforce, i.e. how many of each type of health worker is employed, depends on employers’ (or governments’ or insurers’ or patients’) willingness to pay (wage rate) for each type of health worker. In practice, health care employers may have an idea of the target number of health professionals they require and then decide on wage rates to minimize costs. They will need some knowledge of reservation wages, so jobs are attractive to health workers compared with other competing job opportunities. Demand is therefore a function of trends and policies with respect to broader health care expenditures.

Since wage rates do not reflect consumers’ valuation of the services provided, *how* pay levels are set and regulated is therefore an important part of the institutional design of health worker labor markets (Blau and Kahn 1999; Nickell and Layard 1999). Some regulation of wages and fees is desirable to keep costs down, given extensive market failures. For an organization with a fixed budget, changes in relative wage rates will influence the mix of different types of health worker

that can be employed. Skill mix and task substitution are therefore key issues when deciding on how many health workers to hire. In addition to the level or amount of remuneration, the method of remuneration (e.g. salary, capitation, fee for service, and pay for performance) can influence the demand for different types of labor.

### ***How Are Wages and Reimbursement Levels Set?***

*Summary: Health workers' pay can be determined by market forces or by some form of bilateral pay negotiation (between employers and unions) that may be centralized. How pay is determined influences the flexibility of wages and the extent to which wages help markets clear. Inflexibility can lead to persistent shortages and surpluses of labor. Where employers cannot improve pay to resolve a shortage, they might use terms and conditions of service; or in some cases, additional allowances, such as for rural practice, or for working in a high-cost region. In some low- and middle-income countries (LMICs), these issues are exacerbated if civil service pay is the same across different public sector occupations.*

Some health care labor markets' wages are flexible and determined by market forces, through charging patients fees that are not covered by government or insurer subsidies. Given market failures, this may lead to pay being too high and to equity concerns, as those most in need cannot afford the fees. Some level of subsidy or regulation may be necessary to counteract the effects of these market failures. In LMICs where pay is relatively low, public health workers may also charge informally for their services, with similar consequences.

Health professionals' earnings often constitute the main component of health care costs. Institutional arrangements that govern funding of health care services and price setting play an important role in determining the incentives within the system.

The main issue in pay setting is wage flexibility. This is a key assumption in labor economics: that wages are flexible and adjust freely to changes in demand and supply. If not, equilibrium cannot be achieved and inefficiency persists (box 3.1). However, evidence suggests that where these controls have been introduced, providers adjust volumes of care to compensate for any loss of earnings (McGuire 2000). In other systems where there are no direct price controls and governments and insurers provide a fixed subsidy to patients, increases in subsidies can also lead to increased prices such that only a percentage of the subsidy is passed on to patients (Savage *et al.* 2009).

### ***Centralized Pay***

Pay bargaining arrangements by which employers and trade unions negotiate pay increases and other employment conditions are often used to set pay levels, most commonly in the public sector where trade unions are more dominant (Blau and Kahn 1999).<sup>1</sup> Negotiated wages that are too high encourage a persistent over-supply of labor as wages are extremely hard to readjust, having been fixed through the industrial agreement. Health workers' trade unions, especially for

---

**Box 3.1 Pay Flexibility—The Theory of Compensating Differences**

The theory of compensating (or equalizing) differences is concerned with how the wage rate in a competitive labor market reflects the advantages and disadvantages of jobs (Rosen 1986). This may include the location of the job, flexibility of hours, the type of tasks undertaken, and the types of patients seen. An employer who is offering jobs that have undesirable characteristics (e.g. working in a remote area or in an area with high costs of living) would need to offer higher wages to attract workers, in order to compensate workers for the undesirable characteristics of the job. Employers offering “good” jobs will usually have no problems recruiting, as people are willing to work for lower wages because of the good job characteristics. When individuals choose a job, they therefore trade off wages with other job characteristics.

In this way, the equilibrium market wage captures and reflects all the relative advantages and disadvantages of jobs, again highlighting the central role played by wages in labor economics. Where wages are flexible, there is no need to measure or be concerned with other job characteristics, as these are captured in the market wage rate in efficient labor markets. However, where wages are regulated and inflexible, they may not fully capture the advantages and disadvantages of jobs.

---

physicians, often have strong negotiating power with governments that keep wages relatively high.

Such centralized pay setting reduces wage flexibility. Compared with the private sector, wages will be inflexible and may exhibit little variation across large geographic areas or even whole countries, as is the case for the National Health Service (NHS) in the United Kingdom. An employer facing a shortage cannot raise wages to attract workers, and so relies much more on changing other job characteristics, or face persistent shortages.

Local pay setting, in contrast, say at the level of a region or state or even employer, allows for pay differences between areas that differ in living costs and geographic amenities. In this way, local shortages and recruitment difficulties can be resolved. The empirical measurement of the characteristics of jobs under centralized pay setting then becomes important as these may be the only policy levers available to government and employers to attract and retain workers.

Empirical evidence in general and on health worker labor markets has focused on the impact of pay rigidities in the public sector due to pay bargaining, and their further impact on recruitment and retention problems, the persistence of health worker shortages, and productivity.

For example, assume an industrial agreement in the public sector sets wages for nurses to be the same across geographic areas, but private sector wages for nurses and those occupations that compete for nurses are fully flexible and vary across these same areas. If the public sector wage falls relative to the private sector wage, this will lead to nurses leaving the public sector and working in private sector nursing or other occupations, or in other regions where pay is higher.

This will occur in regions where costs of living are high or the level of amenity is low. In these regions, wages in the private sector are flexible and therefore higher than public sector wages to compensate workers for the high costs of living and disamenity.

Under pay regulation in health care, this type of flexibility is not possible, and has been shown to lead to vacancies and recruitment and retention problems in the NHS (Elliott *et al.* 2007). Further research in England found that pay regulation can also lead to lower productivity overall if highly skilled nurses leave and if the reduction in productivity is greater than the increase in productivity provided by the nurses in their new jobs (Propper and Van Reenen 2010). The gains from pay bargaining of keeping wage costs down might therefore be offset by lower average productivity and quality of care, in addition to persistent health worker shortages.

In practice, it is often desirable that bargaining agreements include adjustments in wages for workers in particular geographic areas or particular types of job. This may include additional allowances for working in a high-cost area, or extra payments for working in rural areas. These allowances may partially offset the impact of wage bargaining on recruitment and retention problems. (The impact of pay differentials across geographic markets will be discussed in more detail in the section on geographic distribution of health workers.)

Centralized pay setting may also be such that the same pay scale is used across different public sector occupations. This creates similar problems of occupational disequilibriums (a relative inability to recruit and retain employees in particular occupations that have more difficult conditions such as long hours, or where there are alternative employment opportunities in the private sector) while also potentially worsening local area disequilibriums.

Such arrangements are quite common in low-income countries' health labor markets, or where health workers are considered civil servants and employed on civil service pay scales. The prevalence of bonus and allowance arrangements in these circumstances suggests that health workers require higher earnings than other types of civil servants to keep them in the public sector. While this helps to resolve the larger problem, it creates others. The fact that such payments are usually flat rate—equal for all points on a pay scale—serves to compress pay scales, with consequences for the ability to retain more skilled or experienced health workers. Recognition of these problems led the Zambian government, for example, to embark on an experiment that sought to “delink” health workers from the civil service, but this ultimately foundered on conflict with the civil service union, which successfully blocked the attempt (Hongoro and Normand 2006; Makasa 2009).

### ***Evidence on Public Sector Pay Issues from Low-Income Countries***

One study addressed in more depth the public sector demand side of the health labor market (Vujicic, Ohiri, and Sparkes 2009). Using case studies of the Dominican Republic, Kenya, Rwanda, and Zambia, it looked at the role of

the government's overall wage bill policy in determining health workforce staffing levels in the public sector. In all four countries, the health sector wage bill is set separately from other types of health expenditure and it is not open to the Ministry of Health to allocate across labor and other inputs at the margin. The demand for health professional labor is therefore generally set by central government, outside the Ministry of Health.

In the Dominican Republic, Kenya, and Rwanda, the health sector accounted for a rising share of the wage bill for periods between 2000 and 2007, reflecting priorities for the health sector, and this trend was also evident in most of a larger set of countries for which similar data were analyzed. However, only in Rwanda of the four case study countries was the overall size of the wage bill maintained over this period: in the others, spending was significantly reduced.

In Kenya, the reduction in spending was judged to have limited the expansion of the health workforce. In Zambia, the situation was difficult to judge owing to concurrent processes to “delink” health staff from the civil service, but the main constraint to health workforce expansion was judged not to be fiscal. Herbst *et al.* (2011) also considered the fiscal space for human resource financing in Zambia and documented its constraints. In Rwanda, the health workforce was successfully expanded (and the stock increased by 30% over a three-year period), and some of the costs for this were transferred to Ministry of Health budgets through a process of performance-based grants (Vujicic, Ohiri, and Sparkes 2009). In the Dominican Republic, expanding the health workforce was not a policy objective, but wage bill restrictions had constrained growth in salary levels with postulated implications for dual practice and hours worked in the public sector.

Because this report, alone among those we identified, documented in some depth the demand side of the public sector labor market, it is possible to reach some tentative conclusions about the state of labor market disequilibrium in each of the four countries. Kenya appears to have had unemployment or a situation of disequilibrium with surplus labor: demand is constrained by macroeconomically driven restrictions on the public sector wage bill as a whole, and pay rates do not adjust downwards to bring the market into equilibrium. Unemployment among qualified health workers is significant. Zambia appears to be in a state of labor shortage: demand existed for health professionals who were not available to fill posts and pay rates did not adjust upwards to bring the market into equilibrium. While the analysis of Herbst *et al.* (2011) tends to support this conclusion for Zambia, it also points to the role played by the demand side in constraining pay rates and other conditions from adjusting upwards—on balance, a conclusion of disequilibrium may require further research on the relative roles of demand (fiscal space) and institutions in constraining wage rate flexibility.

The Dominican Republic appeared to have the most flexible effective pay rates as health workers' unions negotiated reductions in hours to reflect their inability to secure pay increases, bringing into alignment constraints on demand

with reduced supply. The Rwandan health labor market may also exemplify the market clearing model—few health workers are unemployed and the health workforce expanded in line with growing effective demand.

McCoy *et al.* (2008) establish the wide variability in structures and levels of public sector pay (absolutely and relative to living standards and private sector pay) in Sub-Saharan Africa. It is argued that pay levels fell during the structural adjustment period of the 1980s and early 1990s outside Francophone Africa, with data cited to this effect from Tanzania (and more generally for civil servants for 26 of 32 Sub-Saharan African countries between 1986 and 1996), and that concomitantly, nonwage benefits increased proportionately (although the data cited on this point are from Senegal, inside the Francophone group of countries). In the same period, it is argued that wage compression resulted from the award of greater pay increases to lower paid groups alongside the growing importance of flat rate allowances, which would be expected to result in the exit of the most skilled employees at the upper end of the scale to occupations, sectors, and locations where pay scales are less compressed. However, in Ghana, Mozambique, and Uganda, specific policies were introduced to decompress civil service pay scales in the 1990s. This highlights the importance of changes in the distribution of pay and pay scales, and the impact on recruitment and retention.

### ***Health Workers Charging Patients Directly for Services: Price Regulation and Informal Payments***

Some health labor markets leave pay largely to market forces, where health workers can work in private practice and charge fees directly to patients. But with asymmetry of information and the possibility of supplier-induced demand, this approach is likely to result in payments greater than the value of the services provided and may lead to inequality according to need. Patients most in need of health care are usually the poor and disadvantaged. They will not be able to afford the prices charged or will face very large health costs when ill. Insurance and government subsidies of fees, if available, may make services more accessible and so address equity concerns, but do not address the efficiency loss caused by prices (and therefore health workers' earnings) being higher than their actual value to patients and society. To control fees, direct regulation—and therefore negotiation of fee increases between health worker trade unions and governments or insurers—is often used (although as seen, providers may adjust volumes of care to compensate, [McGuire 2000], and higher subsidies may result in only a portion passed on to patients [Savaage *et al.* 2009]).

In low-income countries where there is little pay regulation, or where public sector wages are very low, there may be incentives for health workers to charge informally for the services they provide. McCoy *et al.* (2008) note the importance of private income sources in health workers' incomes. These further complicate the concept of price as some sources, such as informal charges or dual practice income, may be derived from or enhanced by the



health worker's primary employment, while others, such as income earned from agricultural work, are unrelated. The first type of (related) private income sources may be conceptualized as resulting in a divergence of the price paid by the employer and that received by the employee, and indeed this was seen to be among the rationales for the tolerance of dual practice that has been identified in other studies (Eggleston and Bir 2006; González and Macho-Stadler 2013).

Akwataghibe *et al.* (2012) report in detail on the nature of private income sources in two states of Nigeria and find the most significant sources as among those derived or enhanced by public health sector employment: undertaking activities that attract per-diem payments, informal charging, and reselling medicines pilfered from health facilities. Overall, supplementary sources were highly significant: more than 50% of health workers derived more than 50% of their incomes from them.

Gaal, Evetovits, and McKee (2006); Onwujekwe *et al.* (2010); and Hunt (2010) analyze the processes of informal payments in Hungary, Nigeria, and Uganda, respectively. Gaal, Evetovits, and McKee (2006) estimate informal payments in Hungary at 1.5–4.6% of total health expenditure, suggesting a minor role as a source of finance. However, its impact on distribution of financial returns to health workers could be huge: some specialists were estimated to earn 60–236% of their net official income through informal charges in 2001.

In Nigeria, Onwujekwe *et al.* (2010) find that 65% of the expenditure for malaria treatment reported by consumers was not reported by providers. Of particular interest in this paper is that such differences were not confined to public facilities, suggesting that health workers in the private sector may also charge sums beyond those reported to employing institutions, and that the return to employment in private facilities may not be wholly captured by the official salary there too. This finding needs to be interpreted with some caution, however, as the study assumes that all differences between expenditures reported by consumers and providers are informal charges, whereas other explanations, such as a tendency to overstate expenditure by consumers, might apply.

Hunt (2010) finds that informal charges have reemerged in the postformal user-charge situation of Uganda and identifies the existence of informal charges in the private sector too, although at slightly lower levels than in the public sector. However, in the analysis of this paper, private sector “bribes” appear to be predominantly payments by those suffering from respiratory or sexually transmitted diseases, who, according to official policy, should be exempted in the private not-for-profit sector, if they fail to be offered official exemption. This may well not be retained by the employee in the private not-for-profit sector, and may therefore not affect the return to labor supply there.

The analysis of Barr, Lindelow, and Serneels (2009), generated by an experimental game played by Ethiopian nursing students, fails to find an expected reduction in the tendency to expropriate or charge informally, in response to an increased wage rate. Clearly, this can only provide preliminary and provisional evidence on this point.

### ***Skill Mix, Task Substitution, and Elasticity of Demand for Health Workers***

*Summary: With a fixed budget, employers need to decide on what mix of workers to employ in order to achieve their objectives. This will be determined by relative wage rates and the extent to which workers with given sets of skills can undertake the tasks required.*

*Though working in teams is necessary in many health care contexts, there is little empirical literature on the role and nature of teams in influencing health outcomes and costs. Skills may complement each other in undertaking a specific task, or different health workers may be substitutes if they can undertake a specific task but are paid differently.*

*Incentives to change skill mix depend on the extent to which employers have a fixed budget, which is not always the case in the public sector or where defined professional boundaries limit role change. Technological change can also influence the demand for different types of health worker and the value of different sets of skills and how they are combined. Empirical evidence of the determinants and effects of skill mix change is, however, limited.*

Skill mix refers to the potential to substitute tasks and roles between different types of health worker. (Its importance was identified in figure 2.1.) It is also relevant to health workers operating in teams. Some tasks require only one type of labor input, e.g. a doctor, who has a range of skills that are applied to a range of tasks depending on their specialty. A single labor input can also be combined with other inputs, such as a nurse, or capital and technology, such as a drug or procedure. This combination may increase costs and output, and this greater output may be valued more highly if it leads to improved quality of care and patient health status. Or the new input may lower costs by reducing the time it takes to treat a patient (e.g. fewer visits are required), thereby enabling the physician to increase output overall.

For a hospital or physician practice producing a given number of health care services at a specific level of quality, the mix of physicians, nurses, allied health professionals, administrators, and equipment (or technology) will depend in part on the extent to which they work together and combine their skills to undertake specific tasks, which generate a jointly determined output. The skill sets of different labor inputs may overlap, such that doctors and nurses can undertake some of the same tasks, thus creating opportunities for *substitution*. For example, both doctors and nurses can take blood pressure readings, administer drugs, and take blood samples, but in hospitals, nurses undertake these tasks as it is more cost effective for them to do so. Or skills and tasks may be specialized but complementary, for example, a surgeon cannot operate without an anesthetist.

The hospital or physician practice may also have a limited and fixed budget, or have incentives to minimize the costs of producing health care services. The number of full-time equivalent physicians demanded is determined by their wage rate, and the prices of other inputs, which include the wages of nurses, other health professionals, and administrative staff, as well as the cost of capital



and equipment. In the same way, the number of full-time equivalent nurses is determined by their wage rate and the prices paid for other inputs. Recall that the wage rate reflects a worker's value and productivity in a well-functioning market. The most efficient mix of labor inputs (plus equipment) is therefore determined by their relative productivity as reflected by their relative wages. In health care, wages may not reflect the value of labor productivity and so the optimal degree of substitution and complementarity will be determined by the effects of different skill mixes on cost, productivity, and health outcomes.

Two concepts are important here. The first is the elasticity of demand for a health worker, or the extent to which the number of health workers demanded changes as the wage rate changes. If wage rates of doctors rise, the hospital can employ fewer doctors with a fixed budget; if wage rates fall, it can seek to employ more. The second is the cross-price elasticity of demand between inputs. This is a measure of the extent to which different inputs are substitutes or complements. If wages of doctors rise, this influences not only the number of doctors demanded but can also influence the number of nurses and equipment demanded, if doctors and nurses are substitutes or complements. There are, however, no empirical estimates of these elasticities in health care.

To achieve optimal skill mix within an organization, strong incentives are needed to minimize costs, or to maximize the benefits of skill mix change. This may not always be the case in the publicly funded health care sector. Different methods of reimbursement can influence the demand for health workers. Fee-for-service payment is often linked to physicians providing the services themselves, and this provides fewer incentives for the use of nurses, allied health professionals, and physician assistants. Capitation payment—a fixed payment per patient that provides a prospective fixed budget—provides incentives to minimize costs for each patient. Systems with capitation such as the NHS in the United Kingdom and managed care organizations in the United States are more likely to employ nurses and physician assistants and other health workers as they have more flexibility to use the funds in different ways that enable them to minimize costs (Nicholson and Propper 2011).

The substitution of tasks of different cadres of health workers has become central to much health workforce policy. In health care, however, in addition to few incentives for skill mix change, there is much inflexibility in substitution due to regulation and licensing of professions. Clear boundaries between professional groups are enshrined in culture and law. Though nurses may possess the skills to undertake a range of tasks undertaken by doctors, this is often not permitted, meaning that it is illegal in most countries for nurses to prescribe medication, for example. Even within health professions, doctors cannot easily switch between specialties and undertake the tasks of other doctors. Doctors with similar levels of specialty training in one country cannot often practice in another country, without lengthy periods of retraining or supervised practice.

Hospitals as employers often therefore have little flexibility to undertake substitution, as this may require changes to legislation and/or industrial agreements.

This inflexibility leads to persistent inefficiency and again highlights the trade-off between licensing and regulation to maintain minimum quality standards that define clear roles, tasks, and responsibilities, and thus to increases in health care costs due to inefficiency. So although wage rates and earnings may change, this often leads to little change in skill mix in the health sector as doctors and nurses are employed in essentially fixed proportions that change little over time compared with other industries (Nicholson and Propper 2011). There are examples where new cadres of health worker have emerged, such as physician assistants in the United States, though it is less clear whether these are substitutes, complements, or are dealing with previous unmet need and so represent additional costs.

### *Technology and Skill Mix Change*

Technological advancement and its impact on labor inputs are often discussed separately. For example, in health care, the focus is usually on the relationship between different labor inputs, such as doctors and nurses who work in teams. In these analyses, technology is assumed to be exogenous. Given the inflexibility of substitution between different labor inputs, skill mix change and changes in the types of tasks performed by health workers are most often driven by technology-based inputs that influence the demand for labor inputs (Acemoglu and Autor 2011). For example, organizations have a specific set of tasks that need performing to produce outputs, and then choose workers on the basis of the skills that can perform the tasks (Acemoglu and Autor 2011).

Technological change can alter the productivity of all workers across all tasks, or of some workers in specific tasks, or introduce new tasks. There may be different ways to undertake a specific task that use different skills and technologies, and a different set of tasks to produce an output or change total productivity. One reason why we might observe changes in wages rates over time, and in wage differentials between different types of health worker and occupations, is that technological change requires different sets of skills and a changed allocation of tasks. This in turn influences the value of those skills and tasks in terms of changes in wage rates (Acemoglu and Autor 2011). This assumes that wages reflect the value of the output.

### *Evidence of the Effects of Skill Mix Change*

There have been few economic studies of changes in skill mix in health care. Though in labor economics and national labor markets, skill mix change can be analyzed using changes in wage rates (Acemoglu and Autor 2011), we have already established that similar analyses in health care that assume that wages are flexible and reflect the value of output may produce erroneous results. A number of randomized trials and economic evaluations of skill mix change focused on doctor–nurse substitution. The value of output in these studies has been measured using health outcomes and quality of care. Systematic reviews of these studies have found that nurses provide similar levels of quality of care and health outcomes, and for a similar cost, since nurses often have longer consultations than general practitioners (Kernick and Scott 2002; Laurant *et al.* 2004).

This is also the conclusion of studies in LMIC contexts that have considered the planned and resourced reallocation of tasks from higher to lower cadres (in contrast to unplanned ad hoc substitution arising from unfilled vacancies in health teams). For example, comparisons of nonphysician clinicians and doctors in performing obstetric surgery in South Africa, Vietnam, Malawi, and Ethiopia suggested similar postoperative outcomes (Chilopora *et al.* 2007; Gessesew *et al.* 2011; Warriner *et al.* 2006). Similarly, Huicho *et al.* (2008) showed that health workers with shorter training performed at least as well and sometimes substantially better than those with longer training in assessing, classifying, and managing episodes of routine childhood illness and in counseling the children's carers in Bangladesh, Brazil, Tanzania, and Uganda. On the other side, Hounton *et al.* (2009) found significantly higher newborn case fatality rates when cesarean section was performed by clinical officers rather than by obstetricians and general practitioners.

### Supply and Productivity of the Health Workforce

In labor economics, the level of remuneration and the methods by which it is delivered are predicted and can be shown to play a role in influencing labor supply and worker behavior. Applied to health care, financial incentives are expected to influence a range of decisions from occupational choice to retirement to which drug to prescribe, all of which have implications for the number of health workers, the hours they work, and their distribution across countries, specialties, sectors, geographic locations, and health outcomes and health care costs. These behaviors also influence access to health care by the population.

The market failures discussed earlier—particularly wage rates not reflecting consumer preferences, but also supply restrictions to ensure minimum standards and other market imperfections—mean that the expected signals provided by financial incentives are unlikely to help achieve an efficient level of supply or distribution of health workers, as seen in the many countries with persistent shortages of health workers.

The empirical evidence on the impact of financial incentives in health care suggests that they sometimes play a relatively moderate, though statistically significant, role in health workers' decisions. In addition to market failure and other labor market imperfections in health care, this provides another reason why the centrality of wages, earnings, and incentives as policy instruments in labor economics is regarded by some as less useful when applied to health care labor markets.

The key reason is that most health workers do not care only about earnings and consumption, but also about other parts of their utility function, such as patients' health outcomes. Theoretical models of physician behavior in health economics have long recognized this (Ellis and McGuire 1990), and concepts such as intrinsic and prosocial motivation have recognized a much richer set of motivations that drive behavior and that have been applied to health worker behavior by psychologists (Frey and Jegen 2001; Rebitzer and Taylor 2011).

The existence of these other nonpecuniary motivations also makes it important to examine heterogeneity in behavioral responses to incentives, and to capture other nonpecuniary factors that influence decisions and that can be altered by nonwage policies.

### **Education, Training, and Specialty Choice**

*Summary: Economists would argue that the main factor influencing career choices in health care, especially for medical practitioners, is expected lifetime earnings. A focus on earnings in the literature assumes that they capture the “value” to the individual and society of work in that career path or specialty, although there has been little application of this in health economics so far, perhaps because of the weaker link between earnings and health outcomes, productivity, or value to society in health care labor markets.*

*The distribution of doctors across different specialties is a particular issue, given relative shortages of primary care physicians versus specialists. Evidence suggests that the specialty that doctors choose is influenced by differentials in economic returns as well as by other nonpecuniary factors. However, there is much less consensus in the literature on which nonpecuniary factors matter.*

The choice to become a health worker, and subsequent career paths (including specialty choice), is predicted to be influenced by expected lifetime earnings, as well as individuals' skills and other expected job characteristics. Expected lifetime earnings from different occupations and from different levels and types of education and skill acquisition is a central concern of labor and education economics (Card 1999). There is large literature on the “returns to education” and skills, which attempts to identify the effect of different levels and types of education on earnings. Again, a focus on wages is due to an assumption that wages reflect the social returns (i.e. the benefits to society of training and employment) and market value of services provided and so, according to this assumption, there is no need to directly measure the effects of education and skills on productivity or the quality of services provided, as these are reflected in the market wage. These types of analysis also capture the private returns to education (i.e. the benefits to the individual of training and employment) that depend on an individual's characteristics and preferences.

Different lengths of training, graduate versus undergraduate training, and training doctors in rural locations, are all issues that can influence future earnings, the quality of care provided, and other choices such as whether to work in a rural location. There has been some economic analysis on the private returns to different types and lengths of health worker education. Most of this literature has focused on nursing and the move from an apprenticeship model based largely in hospitals, to undergraduate degree training, or comparing registered nurses to less qualified nurses (e.g. enrolled nurses) (Cunich and Whelan 2010; Spetz 2002).

There is little literature on the returns to different types and lengths of education for doctors (Burstein and Cromwell 1985), even though there has been

a move in some countries to shorten undergraduate and specialist training and introduce graduate-entry medical degrees. Most research on the returns to medical education for doctors has focused on the returns to working in different specialties (Cheng *et al.* 2012; Langwell 1982). Specialty choice is a key issue in many countries since there is broad agreement on a long-term imbalance of specialists and primary care physicians. Indeed, the development of primary care as a separate specialty has also been an issue in many LMICs, as well as in the United States. This maps onto a concern about doctors being concentrated in hospitals in urban centers and being reluctant to practice in community-based settings outside major hospitals.

Earnings of specialists are persistently much higher than primary care or community health workers, especially where private practice is the dominant model, and again, medical specialists have much longer periods of training than primary care physicians, constituting barriers to entry. The period of training is determined by medical colleges, which also decide on the number of training places to offer, therefore restricting supply for some specialties, often unrelated to population need.

Differentials in economic returns influence specialty choice. The literature has focused on physicians and the role of future earnings and other factors influencing such choice (Nicholson 2002; Sivey *et al.* 2012). These studies show that although earnings do play a role in specialty choice, so do other factors such as medical education, debt, and flexibility of hours. However, there is much less consensus in the literature on the role of nonpecuniary factors.

Carnoy *et al.* (2012) estimate returns to higher education in the BRIC countries (Brazil, the Russian Federation, India, and China) and find large differences between them. In Russia between 1998 and 2001, private rates of return to medical education were negligible for men but comparable to the returns to higher education in law, engineering, and humanities for women. Higher returns to women reflect the poor job opportunities for women without higher education relative to men. Private rates of return declined over the period as tuition fees increased. In Brazil, private rates of return to medical education were high, but not as high as other higher education investments. Medical education was not considered separately from other higher education in China and India.

Our search did not find studies calculating the rates of return to any type of health professional training in low-income countries, though some recent qualitative work suggests that rates of return may be limited by shortcomings in training institutions (Celletti *et al.* 2011; Ferrinho *et al.* 2011).

### **Workforce Participation and Hours Worked**

*Summary: Higher wages are assumed to provide an incentive to work more hours and to encourage workers currently not working to choose to work. Empirical evidence suggests that changes in the level of pay influences hours worked and workforce participation. There is also the possibility and some evidence of doctors on relatively*

*high incomes working less, as they prefer to take more leisure time when earnings increase.*

Examining how changes in wages influence hours worked and labor force participation is a central concern of labor economics (Blundell and Macurdy 1999). Structural labor supply models assume that workers, in determining the number of hours they work, are motivated by their income and leisure. In labor economics, these issues are driven by policy changes in taxation and welfare benefits that influence wage rates and then labor supply. A key concept is the elasticity of labor supply, which measures the responsiveness of hours worked to changes in wages. Higher wages are assumed to provide an incentive to work more hours and to encourage workers currently not working to choose to work. Often, a focus of studies examining the effect of wages on labor supply is on the labor supply of women and the role of children in their preferences for income and leisure. Models also involve dynamic and intertemporal labor supply decisions, where individuals attempt to make decisions about hours worked over their lifetime.

There is also the possibility of a “backward-bending” labor supply and a negative labor supply elasticity by which an increase in wages causes a fall, rather than an increase, in hours worked. This happens because wage changes affect income and the rate of substitution between leisure and income. Under normal conditions, the substitution effect dominates the income effect, such that a wage rise leads to the marginal utility from the additional income from working an extra hour being greater than the utility from an extra hour of leisure—an increase in wages therefore leads to an increase in hours worked. If the income effect dominates the substitution effect, the increase in income is sufficiently large that the worker has a higher utility from an extra hour of leisure than from an extra of work, and so hours worked fall.

Labor supply elasticities are therefore important pieces of information that help assess the effectiveness of changes in wages and translate them into changes in hours worked, and then into changes in the number of health workers employed.

In health care, labor supply models have been estimated from a policy concern about the general impact of changes in wages on the labor supply of doctors and nurses. The driving policy concern has therefore not been about taxes or welfare benefits, but about the impact of changes in regulated wages and fees that are often the subject of pay negotiations between government, insurers, and health workers’ professional organizations.

Changes in the level of pay have been shown to influence hours worked and workforce participation of doctors and nurses (Antonazzo *et al.* 2003; Baltagi, Bratberg, and Holmås 2005; Hanel, Kalb, and Scott 2012; Sæther 2005; Shields 2004). The impact of higher hourly earnings is to increase hours worked and workforce participation, with this effect being relatively small, but usually statistically significant, in most labor supply models for both doctors and nurses. There is little strong evidence of backward-bending labor supply: weak evidence can be observed in aggregate data in several countries where the



average hours worked and the number of patients seen by doctors are falling while remuneration levels and prices are increasing (Scott 2006; Whalley, Gravelle, and Sibbald 2008).

Qin, Li, and Hsieh (2013) use census data to estimate the labor supply response (weekly working hours) of health professionals to changes in the hourly wage rate when considering policies of increasing wages or increasing training opportunities as alternative responses to rising demand for services in China. Their analysis contrasts self-employed health professionals (who have relative flexibility in setting their working hours) and employees (who are much more constrained). Their results suggest supply elasticities in response to the hourly wage rate of 0.575 for self-employed health professionals (for a 1% increase in the wage rate, a 0.575% increase in numbers of hours worked) and 0.02 for employees (virtually no supply response). The authors attribute the difference to constraints on employees' ability to increase and reduce hours worked at the margin.

Qin, Li, and Hsieh (2013) note that their result for self-employed health professionals constitutes a high degree of elasticity relative to estimates for countries in the Organisation for Economic Co-operation and Development (OECD), and one that may underestimate the longer-term response of reducing constraints to labor supply adjustment. They conclude that greater upward flexibility of health professional pay, which is constrained by regulation in the Chinese health system, may be a more cost-effective policy option than investment in health professional education as a means of increasing labor supply, and that it may slow health price inflation, which they assume reflects the increasing shortage of health professionals in the system.

### **Migration and Retention**

*Summary: Wage differentials influence retention of health workers, both in specific jobs and in relation to out-migration. Higher wages reduce attrition and induce health workers who would choose to migrate at current salary levels to remain in their jobs. However, this relationship becomes more inelastic at higher levels of salary increases, suggesting that policy tools aimed at enhancing recruitment and retention of health workers should not be restricted to salary improvements alone. Job attributes equally play an important role with no single job attribute dominating. Empirical evidence provides a mixed picture of the impact of wage differentials on migration.*

### **Migration**

Using payroll data from the government of Ghana, Antwi and Phillips (2013) find elasticities of attrition from the public sector payroll of around  $-0.1$  for younger health workers identified as potential migrants (a 10% increase in wages reduces annual attrition by 1% from its 8% base level), but insignificant or marginally significant effects for other groups. They suggest that their results indicate that the effect of wage rates on retention operates largely through migration.

Using a different approach (contingent valuation), Serra *et al.* (2010) find a fairly consistent response over a range of 40–300% salary increases, namely that each increase of about 40% of the current salary level induces a further approximately 5% of nurses (who would otherwise have migrated at current salary levels) to stay in Ethiopia. Responses are highly inelastic to further salary increases beyond 300%, however, with 30% of nurses choosing to leave Ethiopia if given the chance, irrespective of Ethiopian salary levels (Serra *et al.* 2010, 42, figure 6.1). For doctors, salary increase increments of about 30% of the current salary level each induces 3–4% of doctors to stay (who would otherwise have left in the near term), but the relationship becomes more inelastic to further salary increases above 240% of the current level, and almost completely inelastic to further salary increments above 500%: at that point, 85–90% of doctors would choose to stay in Ethiopia (Serra *et al.* 2010, 43, figure 6.2).

In contrast, Vujicic *et al.* (2004) in a survey of health professionals in six African countries fail to find a relationship for wage differentials between migrant origin and destination countries, on the one hand, and the prevalence of intention to migrate, on the other. They suggest that the universally large gap in wages between migrant origin and destination countries may explain why relatively small differences in origin countries' wage levels have limited explanatory power, and advise that countries aiming to limit migration may need to focus on nonwage instruments. This study did not disaggregate health professionals by age and internationally scarce skills as Antwi and Phillips (2013) did, and may consequently have missed finding the relationship in the most likely affected subgroup.

International recruitment policies and health labor market forces have played an important role in fueling the out-migration of health workers from LMICs (WHO 2006). Changes in population demographics as well as in international labor market forces and economic conditions have partly contributed to a growing demand for internationally trained health workers in high-income countries (Pond and McPake 2006). For example, in the United Kingdom, the sharp increase in health expenditure of the NHS in 1998 resulted in a rise in the demand for internationally trained health workers. In the United States, capacity restrictions in U.S. medical schools after the Balanced Budget Act of 1997 resulted in a fall in the supply of domestically trained medical doctors, with a resulting increase in absorption of internationally trained health workers (Pond and McPake 2006).

To meet this growing demand, high-income countries have used various strategies for recruiting international health workers, including relaxing immigration policies and licensing requirements (Bach 2007; Birrell 2004; OECD 2002; Runnels, Labonte, and Packer 2011). For example, the U.K. government in 2001 embarked on an advertising campaign to recruit internationally trained doctors and nurses; by 2003, the number of doctors and nurses had grown annually by around 21% and 20%, respectively, with international recruitment accounting for 80% of the growth in the number of doctors and 73% of the growth in the number of nurses. Of the additional overseas-trained physicians



registered during 2002 and 2003, 24% came from Sub-Saharan Africa (Pond and McPake 2006).

Recognizing the adverse impact of their international recruitment policies on the health workforce of LMICs already experiencing shortages of qualified health workers, high-income countries adopted international recruitment codes of practice and bilateral agreements to limit such recruitment (Kurowski *et al.* 2012; Runnels, Labonte, and Packer 2011; WHO 2010). However, the impacts of these policies remain unclear, partly because of the difficulty in identifying the impact of a code among other policies (Buchan *et al.* 2009). For example, although the U.K. Department of Health in 2001 set out guidelines restraining the active recruitment of health workers from 150 countries, including countries in Africa and the Caribbean (despite the concurrent advertising campaign mentioned above) (Department of Health 2001, 2004), the number of Caribbean-trained nurses registered to practice in the United Kingdom grew by 40% between 2001 and 2005 (World Bank 2009) while the number of Sub-Saharan Africa-trained doctors registered to practice in the United Kingdom tripled between 2001 and 2003 (Pond and McPake 2006). However, by 2006, international recruitment had declined in the United Kingdom, partly reflecting changes in general immigration policies and in professional licensing requirements for professional health workers entering the country (Buchan *et al.* 2009).

### **Retention**

Three studies were undertaken through the Regional Network for Equity in Health in East and Southern Africa (EQUINET) in Tanzania, Swaziland, and Kenya. The first two focused on nonfinancial incentives; the third considered financial incentives as well.

In Tanzania, Munga and Mbilinyi (2008) found that nonfinancial incentives have the potential to significantly affect health workers' willingness to stay in their posts but that ineffective implementation of such incentives limits their practical impact. Examples included the lack of transparency in allocating promotions and training opportunities, and the inadequate provision of benefits, such as stipulated housing entitlements. Masango, Gathu, and Sibandze (2008), in Swaziland, found nonfinancial factors positively associated with retention: job satisfaction, a perception of equality of treatment by the employer, the discretion that health workers felt they had in carrying out their role, and the opportunity that role gave to help others.

Ndeti, Khasakhala, and Omolo (2008) also found nonfinancial incentives highly valued in Kenya, and documented the disparity of financial and nonfinancial incentives between rural and urban areas. In part, this disparity stemmed from the ability of better organized facilities, often in higher income areas, to adopt measures introduced to retain health workers.

In Senegal, Rouleau *et al.* (2012) found some correlation between elements of a job satisfaction survey and job search activity and turnover, using a longitudinal study of midwives. In the West Nile region of Uganda, Onzubo (2007) found low

salary to be overall the most frequently estimated reason for attrition by health staff who remained in public and private not-for-profit hospitals, although in public hospitals, political and managerial explanations were more frequently offered.

Penn-Kekana *et al.* (2005) (for three provinces in South Africa with high vacancy levels) and Mangham and Hanson (2008) (for Malawi) used discrete choice experiments (DCEs) to evaluate the role of different job attributes in retention and related human resources outcomes in the health sector. Both studies found that no single job attribute dominated employment preferences of South African nursing and midwifery staff or Malawian public sector-registered nurses, and that almost all were willing to trade among attributes. In South Africa, good management and a well-equipped hospital were as important to nurses as a 15% pay increase. In Malawi, opportunities to upgrade qualifications, provision of government housing, and pay had the greatest impact on the utility associated with a particular job. Mangham and Hanson (2008) conclude that a range of policy instruments is available to influence Malawian nurses' motivation, recruitment, and retention, beyond salary increases.

In Vietnam, Witter *et al.* (2011) found four "directions of travel" over the course of health professionals' careers, militating against the retention of health professionals in public, rural, preventive, and primary service provision, motivated mainly by financial incentives.

### **Dual Practice**

*Summary: The dual practice literature consists almost entirely of two types of material: descriptive statistics and theoretical analysis. While some theoretical studies suggest that allowing dual practice might increase social welfare and improve quality in the public sector, others suggest that quality might fall in the public sector as physicians skimp on hours or have incentives to redirect profitable patients to their private practice. However, these theoretical models have yet to be tested empirically.*

The allocation of resources across public and private health sectors is a contentious issue in most countries, including the extent to which doctors and other health workers undertake work in both sectors. Standard approaches in labor economics do not address these issues directly, as they assume that workers choose the highest paying job and that employers discourage their workers from supplying their labor to other firms (Eggleston and Bir 2006).

The dual practice literature consists almost entirely of two types of material: descriptive data and theoretical analysis, with very limited use of theory to analyze data or use of data to test or generate theory. Socha and Bech (2011) provide a review of the theoretical literature and highlight the need to test the assumptions that drive it, such as income maximization, as it seems to fit poorly the descriptive empirical observation that exists. Almost all the literature considers only dual practice among doctors; evidence on other cadres such as midwives, who have also been identified as dual practitioners, is almost entirely absent.

Eggleston and Bir (2006) review five models of dual practice. Most are consistent with either positive or negative relationships between dual practice, social welfare, and quality in the public sector, implying a need for more empirical evidence.

Eggleston and Bir (2006) also provide some analysis of the fit of theoretical models with the limited empirical evidence. For example, they argue that documented government attitudes to dual practice support the theoretical idea that allowing dual practice enables government to recruit quality providers at modest budgetary expense, and other evidence (including some from Bangladesh) supports the argument that dual practice has the greatest value to more specialist health professionals. They cite a World Health Organization study (Hicks and Adams, 2001) to support a postulated negative correlation between public salary levels and prevalence of parallel income-generating activities by health professionals, including dual practice. Nevertheless, Eggleston and Bir (2006) find very limited empirical evidence to compare with theoretical predictions, searching even further back in the literature (to around 1990) than we have (to 2006).

More recent reviews of the broader literature confirm that there is much descriptive analysis and debate, but hardly any empirical studies that examine the impact of changes in dual practice on health care costs and health outcomes (González and Macho-Stadler 2013; Socha and Bech 2011). González and Macho-Stadler (2013) argue that there is little consensus in the literature on the net effects of dual practice on costs, quality of care, health outcomes, and access to care. This is reflected in a wide variety of arrangements to regulate dual practice, either through bans on it, improved contracts for public sector employment, or limits on the amount of hours or earnings from private practice. The authors note that these regulations are most prevalent in high-income countries, with dual practice largely unregulated in many LMICs.

The theoretical models reviewed by Eggleston and Bir (2006) and the more recent discussion of González and Macho-Stadler (2013) tend, however, to ignore the consequences of market failure in health care. They assume that physicians in private practice have higher levels of skill, are more productive, and provide care that is more highly valued than physicians in the public sector—presumably because they have higher wages than those in the public sector, and this would be the case in a well-functioning labor market. Under these assumptions, banning or restricting dual practice leads to the most productive physicians moving to the private sector, and these effects may be worse in low-income countries, given the larger gap between the sectors (González and Macho-Stadler 2013).

However, these conclusions do not hold if the link between earnings and the value of services provided is equally weak in both sectors because of market failure. Though physicians may be more productive in some cases, this may not translate into their being more skilled—it may rely on better resource availability. In addition to market failure, these models ignore other motivations such as higher-skill physicians caring more about patients' health status than income, and therefore preferring to spend more time in the public sector where, typically, patients have a higher need for health care services.

The public sector typically has less ability to increase wages to attract physicians from the private sector, and so wages could remain low compared with those in the private sector, creating persistent disequilibrium and recruitment problems in the public sector. The size of these problems will depend on the extent of the between-sector differences (González and Macho-Stadler 2013), which are likely to be larger in low-income than high-income countries. As we have seen, public sector wages may be centrally determined or part of a civil service pay scale. There may be little scope for pay to reflect performance, and little opportunity to progress up the salary scale. The private sector will usually be based on fee-for-service payment where the volume of care is related to income, thereby providing incentives to see more patients and to have higher productivity levels than in the public sector.

The public sector may therefore focus on nonwage aspects of jobs to attract health workers from the private sector, including offering the opportunity to undertake private sector work, the amount of which could be regulated to maintain time spent in the public sector. But the public sector could also offer some doctors the intellectual challenge of academic work that can enhance their reputations, and allow them to work with different groups of patients. Doctors may though prefer private sector work because they have more autonomy.

Conditional on the regulations on dual practice, the issue then becomes whether the allocation of time between sectors can be changed to improve population health at least cost. Long waiting lists and times at public sector facilities, or a high need for health care among groups that do not currently access health care, may be evidence of a shortage of doctors in the public sector, especially when private sector waiting times and lists are short or those patients being treated in the private sector are more healthy and less in need of health care.

The need to encourage doctors to allocate more time to the public sector therefore becomes an important objective of government policy. A doctor who transfers a session (say, a half-day) of work from the private to the public sector will improve the health of patients if the marginal health gain of the additional session in the public sector is greater than the health gain given up in the private sector. In a well-functioning market, this would be reflected by a higher hourly wage in the public sector. But of course in health care, wages are not related to marginal health gains in either the public or private sector, and so if doctors care more about income or other factors than patients' health they will not reallocate their time even though health gains are larger in the public sector. As said, design of effective policy interventions to achieve this requires an understanding of the factors that can potentially influence doctors' decisions about how they allocate time across the sectors.

What is clear from the González and Macho-Stadler (2013) model is that policies that ban or restrict dual practice could potentially lead to doctors leaving or spending less time in the public sector. However, how much they do that depends on how much they are motivated by income or other attractions of private practice, the transactions costs of moving, and the differences in job characteristics between the two sectors. If the doctors who

leave produce greater health gains in the private sector than they did in the public sector, these regulatory policies will not improve efficiency in the public sector, but population health may be higher from a societal point of view.

Some studies report significant levels of dual practice in various settings. Both Gruen *et al.* (2002) (Bangladesh) and Jumpa, Jan, and Mills (2007) (Lima, Peru) document beliefs and attitudes among dual practitioners; Jumpa, Jan, and Mills (2007) find income generation the prevalent explanation, although they identify other benefits including skills development and greater exercise of clinical autonomy. Gruen *et al.* (2002) estimate incomes earned and consultation fees and find the latter significantly associated with the degree of specialization, surgery or hospital admissions, and employment of own staff, while 17% of the sampled doctors reported earning between two and four times as much in the private as the government sector. Predictors of total income levels included age, time in practice, employment of own staff, and (negatively) employment in a primary health care facility.

Russo *et al.* (2013) (Praia, Cape Verde; Bissau, Guinea-Bissau; and Maputo, Mozambique) evaluate the factors associated with the decision to engage in dual practice and find significant associations with years as a physician, having a specialization, age, and period working outside the capital city.

Both Jumpa, Jan, and Mills (2007) and Russo *et al.* (2013) identify categorization problems in different forms of private sector activity undertaken by public sector doctors, and the phenomenon of private sector activity within public sector premises. Russo *et al.* (2013) conclude that where this option is available, it appears to limit dual practice in the form of private practice outside public sector premises, and is perhaps preferred because it removes the need to invest in separate private clinics.

Two articles focus on regulation of dual practice. González and Macho-Stadler (2013) provide a theoretical framework considering the implications of different regulations. The main conclusion is that limiting dual practice will be more effective if the focus is on limiting earnings rather than limiting activity, but as discussed above, this conclusion is contingent on the normal operation of market forces—in this case, that the private market is able to recognize and reward ability. Garcia-Prado and Gonzalez (2007) document the experience of different types of regulation in a range of settings, and conclude that implementation of regulations poses such difficulties that it is impossible to evaluate impact. Successful regulatory experience is only documented in high-income countries, although as discussed, encouraging the development of private practice in public facilities is also prevalent in low-income countries, but with unknown impact.

### ***Geographic Distribution of Health Workers within Countries***

*Summary: Shortages in areas of high need for health care, including rural areas and areas that include socially and economically disadvantaged populations, indicate that local labor markets in health care do not operate efficiently. A compensating wage*

*differential approach highlights the role of the relative advantages and disadvantages of local areas.*

*Many countries have a range of policies and programs to encourage doctors to locate and practice, even temporarily, in underserved remote and rural areas. These include financial and nonfinancial incentives, restrictions on mobility such as mandatory practice in rural areas for immigrant health workers, bonded training schemes, and restrictions on entry to overserved areas.*

*Despite a considerable literature identifying factors that influence the recruitment and retention of doctors in remote and rural areas, to date there exists little rigorous evidence about which policies or programs are the most effective in increasing the supply of doctors to underserved areas, and more specifically the amount of incentive required to encourage doctors to move.*

In geographically defined labor markets, whether a country or regions within a country, differences in earnings and other job characteristics as well as characteristics of geographic areas should in theory influence the mobility of workers between areas. Where wages are flexible, reflect the value of services provided, and there are no transaction costs of moving, a shortage in one area will push up wages offered in that area relative to other areas. This will lead to workers moving into that area until the wage falls back to spatial equilibrium.

In labor economics, interest in local labor markets is driven by attempts to explain differences in productivity, wages, and employment across spatial labor markets, and how changes in demand and supply conditions in local labor markets, or in the aggregate, have differential effects across local areas (Moretti 2011). In health care, interest in geographic distribution is driven by the perception of persistent shortages that lead to concerns of equity of access to health care in addition to concerns about efficiency. Shortages in areas of high need, including rural areas and areas that include socially and economically disadvantaged populations, indicate that local labor markets in health care do not operate efficiently. A compensating wage differential approach highlights the role of the relative advantages and disadvantages of local areas. For example, in rural areas there are typically longer working hours, higher levels of being on call, and the need to work in relative professional and social isolation.

A previous section on pay setting highlighted how centralized wage bargaining can lead to persistent shortages. However, even where wages are flexible, in rural areas, diseconomies of scale lead to high unit costs of production and make it less profitable for firms to locate there, which may be efficient from the firm's perspective, but in health care, a concern for equity of access will lead to arguments for government intervention. There are two key reasons for government intervention: to compensate workers for the disadvantages of jobs and areas, and to subsidize the provision of health services that have high unit costs because of diseconomies of scale.

Many countries have a range of policies and programs to encourage doctors to locate and practice, even temporarily, in underserved remote and rural areas. These include restrictions on mobility such as mandatory practice in rural areas for immigrant health workers, bonded training schemes, or regulations restricting



entry to overserved areas. Financial or nonfinancial incentives are also used in many countries to encourage health workers to move to underserved areas and to stay in those areas. For example, short-term contract-based employment and high-remuneration packages have been used as policy tools in Turkey and Indonesia to attract health workers to rural areas (Rokx *et al.* 2010; Vujicic, Sparkes, and Mollahaliloglu 2009).

Imbalance in distribution of clinical schools has been highlighted as a determinant of the inequitable rural–urban distribution of health workers; the establishment of rural clinical schools may prove an important policy tool (Raha, Berman, and Bhatnagar 2009a). Other recent policies include exposure to rural practice during training and targeted recruitment of health workers who grew up in rural areas. Government intervention may also attempt to reduce the transactions costs of moving. These include family factors such as schooling and employment for a partner, and assistance with the costs of establishing a small business.

While different combinations of financial and nonfinancial incentives have been used to attract health workers to rural posts, lower economic conditions and standards of living prevailing in rural areas represent important challenges for retaining health workers there (Rao *et al.* 2010; Rokx *et al.* 2010). This raises questions about what types of incentives or combination of them will compensate health workers for the loss of utility associated not only with the characteristics of rural jobs but also with the socioeconomic characteristics of rural areas. Despite a wide literature identifying factors influencing recruitment and retention of doctors in remote and rural areas, to date there exists little rigorous evidence about which incentive policies or programs are the most effective in increasing the supply of doctors to underserved areas, and more specifically the amount of incentive required to encourage enough of them to move (Barnighausen and Bloom 2009; Buykx *et al.* 2010; Grobler *et al.* 2009). These systematic reviews found very weak evidence of effectiveness (but many poor study designs).

Lemiere *et al.* (2011) make the relatively neglected point that a key factor in the underrepresentation of health staff in rural areas is low effective demand there. Low incomes and consequent lower willingness and ability to pay restrict the profitability of for-profit providers, the scope for sustainability of not-for-profit providers, and the potential for fee income of public providers. In the public sector, fiscal centralization restricts the funding for public sector services. Supply of health professionals to rural areas, or the relative willingness to work there, is argued to be differentiated by type of health professional and other health worker characteristics such as gender and family circumstances, but to be on balance lower than the supply to urban areas, even with higher compensation.

However, evidence from Ethiopia (World Bank 2008) suggests that in practice, monetary compensation to health workers is higher in urban areas, consistent with the arguments of Ndeti, Khasakhala, and Omolo (2008) (for Kenya) and Witter *et al.* (2011) and Vujicic *et al.* (2011) (for Vietnam). Serra *et al.* (2010) find that doctors in Ethiopia earn more in rural areas than in urban areas,

and nurses only slightly (and statistically insignificantly) more in urban areas. It is unclear why the evidence of World Bank (2008) and Serra *et al.* (2010) for the same country at a similar time is contradictory.

Designing incentives to encourage doctors to locate and remain in remote and rural areas requires an understanding of the various factors that motivate their location decisions. In the absence of data on revealed preferences, DCEs are increasingly used. While Penn-Kekana *et al.* (2005) and Mangham and Hanson (2008) used DCEs to address the issues of motivation and retention, others have used the same approach to understand the opportunities and costs of incentivizing health workers to rural jobs, and the results of the earlier reported studies could be used to assess alternative strategies for incentivizing them in this way.

Vujicic, Alfano, and Shengelia (2010), for example, found that while the level of pay was the most important job attribute for nurses in Liberia, opportunities for further study were most important for doctors in Vietnam. Translating this to a cost-effectiveness analysis confirmed that the most cost-effective strategy to increase recruitment in rural areas was increasing pay in Liberia and improving opportunities for further study in Vietnam. The Liberian and Vietnamese studies derived the list of attributes used in the DCE by prior qualitative work (Attah *et al.* 2010; Witter *et al.* 2011).

Lagarde, Blaauw, and Cairns (2012) took a similar approach in South Africa, first using DCEs to understand the elasticities of nurses' labor supply to rural areas in response to policy controlled variables, and then calculating the cost-effectiveness of the policies. Their analysis found that the most cost-effective intervention was to select nursing students more likely to accept rural posts—an option not considered by Vujicic, Alfano, and Shengelia (2010)—followed by an offer of preferential access to specialist training. Vujicic *et al.* (2011) further analyzed the differences in values of job attributes expressed by those who were (or were not) looking for a new job, had lived in a rural or urban area prior to starting medical studies, and whose parents had (or had not) graduated with a university degree. The lowest values for an urban job location were expressed by those who were looking for a new job, originated from a rural area, and whose parents had not graduated from university.

Kruk *et al.* (2010) used a DCE to explore the job attributes that would persuade Ghanaian medical students to accept a rural post. Improved infrastructure, supportive management, and a 100% salary bonus were the most important predictors, and all three combined predicted that 90% of respondents would choose a rural post.

Lori *et al.* (2012) found dominance of training opportunities for midwifery students in Ghana in choosing between urban and rural work alternatives, and Appiah-Denkyira *et al.* (2013) report that for three-year diploma nurses, a fixed term of two years followed by study leave was the dominant job attribute predicting choice of a rural post. Similar results have been reported by Raha, Berman, and Bhatnagar (2009b) in India. Using semi-structured interviews, they showed that compared with other nonfinancial incentives such as increased



training opportunities, good housing, and faster promotions, the most attractive incentive was a 50% reservation in a postgraduate course on completion of the rural posting: over 80% of undergraduate medical students who would otherwise have pursued a postgraduate degree on graduation were willing to accept a two- or three-year rural posting if offered that. Similar results were observed with nursing students.

Hanson and Jack (2010) report results of a DCE among doctors and nurses in Ethiopia, finding that nurses are more responsive to pay levels than doctors, but relative to pay, place a higher value on working in Addis Ababa relative to alternatives than doctors do. There are, however, difficulties in interpreting these results as the alternatives offered were not identical: nurses chose between Addis Ababa and a rural location that could be quite remote, whereas doctors chose between the capital and a regional town. Furthermore, pay responses were relative to current pay levels, which were of course lower for nurses than for doctors. Further modeling of the factors influencing the values stated found significant roles for marital status, sex, children, current location, current housing, and interactions between these variables.

Comparing the DCE studies described in the preceding section suggests the following: that generalization about the relative importance of financial and nonfinancial incentives is unwarranted; that financial and training-related incentives may prove most important in many settings—although training-related incentives were not included in the study by Hanson and Jack (2010); and perhaps that financial incentives are likely to be more important in the poorest settings (Ethiopia and Liberia rather than the middle-income context of Vietnam) and for lower cadres (nurses rather than doctors). The latter two suggestions provide hypotheses for comparison with further evidence.

The two incentives that are often important (financial and training incentives) were found to be juxtaposed in Ethiopia by Serra *et al.* (2010). For doctors, pay was found to be higher in urban areas while training opportunities were more constrained. For nurses, the reverse was the case with rural pay similar or slightly lower in rural areas but training opportunities greater. Serra *et al.* (2010) also used a contingent valuation approach, similar to a DCE, but focused only on determining the salary at which nurses and doctors would accept a post in a rural area. To attract 80% of nurses and 65% of doctors to a rural post, salary levels of 284% and 245% of the current levels, respectively, were estimated to be required.

### **Productivity and Performance**

*Summary: Productivity and performance can be influenced by monetary and nonmonetary rewards. Most evidence concerns the implications of different forms and structures of monetary reward. Payments can be made to the health provider in exchange for working for a specified time period (e.g. salary, sessional payment); for providing specific services, treatments, and episodes (fee for service); for providing care for a patient or specific population (capitation); and for providing a pre-specified level of quality of care or health outcomes (e.g. performance-based pay). Payments are also*

*characterized by different distributions of risk between health care provider and employer or funder.*

*“Pay for performance” programs are attempts to reform payment structures so as to improve productivity and share risk. Types of payment mechanism perceived as low risk and low incentive in practice use a combination of financial and nonfinancial rewards over the longer term.*

### ***The Impact of Remuneration on Productivity and Performance***

The existence of a principal–agent relationship between doctors and patients means that much of the literature in health economics is focused on designing and evaluating optimal payment mechanisms for health workers, mainly physicians (Malcomson 1999; Prendergast 1999). The mechanism needs not only to ensure the efficient supply of hours, but also to provide incentives for increased performance, from the agent, which aligns with the principal’s objectives. Designing incentive-compatible contracts requires that the principal’s objectives can be met, while ensuring that the contract is attractive enough to agents. These contracts depend on the extent to which performance and productivity can be observed.

A key part of this literature is multitasking, which recognizes that complex jobs such as medicine may not be suited to strong performance-incentive measures. This is due to the wide range of different tasks and the inability, or sometimes high costs, of measuring performance along all dimensions. There may be unintended consequences of performance monitoring and measurement, such that it is difficult to assess if efficiency overall has improved (Holmstrom and Milgrom 1991). With such multitasking, performance incentives for measurable activities may divert efforts from equally important but unmeasurable aspects of performance. In medicine, which already has a high level of intrinsic motivation to improve health outcomes, the need for such strong incentives is diminished (Mooney and Ryan 1993).

In addition to the incentives concerning the level of wages that are the focus of traditional labor economic models, incentive contracts that encourage effort and performance focus on different types and methods of remuneration and employment contracts. These are also more suitable in health care environments, where many health workers may not be paid a salary and are self-employed in private practice. Self-employment coupled with fee-for-service payment, capitation payment, or pay for performance may all have different effects, not only on hours worked but also on the allocation of those hours across different treatments and types of patient, as well as effects on health outcomes and costs. Though fee for service, salary, capitation payment, and pay for performance are the main methods of payment, they can also differ depending on other characteristics of the payments (Scott *et al.* 2011).

- Payments can be made in exchange for the following provider behaviors: working for a specified time period (e.g. salary, session payment); providing specific services, treatments, episodes (fee for service); providing care for a patient or specific population (capitation); and providing a prespecified

level, or increases in the level, of quality of care or health outcomes (e.g. performance-based pay).

- Payments may be linear or nonlinear: linear, so the same payment is made for each additional unit of payment (e.g. service provided); or nonlinear, such that payment is conditional on reaching a threshold or target, or a series of thresholds, or that the amount of the payment changes with each additional service provided.

Payments may also be risk adjusted. Capitation payments provide incentives to minimize costs, which provide incentives to select only the healthiest patients to treat. Capitation payments can therefore be risk adjusted to avoid providers choosing less complex patients to treat (cream skimming), which is an unintended consequence of this form of payment. Risk adjustment often weights the payments per patient for the expected costs of treating that patient. For example, elderly patients may attract higher payments than younger patients.

Payments can also be based on administrative rules, such as seniority. This is likely to be reflected in salary payment with automatic annual progression between increments on the scale to reflect accumulated experience. This may be relatively efficient where performance cannot be easily observed or measured (Prendergast 1999).

The timing of payments may be in advance (prospective payment that provides a fixed overall budget) or once performance can be measured and payment adjusted. Such retrospective payment can be capped or unlimited with respect to the total payments that can be made. Once the cap is reached, either no further payments are made or the amount of the unit payment is reduced.

Another key feature of payment systems is the extent of risk sharing between providers and employers or funders (Ellis and McGuire 1990). Uncapped fee for service means that patients or funders are exposed to the full financial risk of any change in the volume of care provided. In this setting, providers have few incentives to control costs. At the other extreme, capitation payment shifts the financial risks largely to the provider, who must absorb the additional costs of patients with complex and expensive health problems. Incentives for cost control by providers are high. The degree of risk sharing influences the strength of the incentives.

A combination of the above approaches is often regarded as efficient as it avoids the extreme incentives of each single method (Robinson 2001). Pay for performance is usually added to an existing method such as salary payment or fee for service. However, the efficient mix of different methods of payment depends on the context.

There is also a large literature examining changes in the method by which health professionals—mainly doctors—are paid. Cochrane systematic reviews have all found that different methods of payment influence clinical behavior and the quality of health care, although the effects are small and the evidence is mixed and of variable quality (Eccles *et al.* 2010; Gosden *et al.* 2000, 2001; Scott *et al.* 2011).

Compared with salaried and capitation payment, fee for service encourages a higher volume of care provided. There is an emerging consensus that such payment does not encourage optimal care for patients with chronic disease, and there are payment models introducing blended payments that include a capitation payment and an element of pay for performance as part of “medical homes” and accountable care organizations in the United States (Merrell and Berenson 2010). These have existed for some time in the United Kingdom for general practitioners who received 25% of their earnings through the Quality and Outcomes Framework pay for performance system (Roland 2004). Such models require standardized measures of performance that are only possible with other interventions, including information technology infrastructure and agreed guidelines on treatment options for specific diseases.

The reported doubts about the effectiveness of systems such as pay for performance are concerned not only with the poor methodological design of the studies themselves, but also with the poor design of the systems (Campbell *et al.* 2010). Avoiding unintended and undesirable consequences—there may also be some unintended but desirable consequences (Sutton *et al.* 2009)—can be partly achieved through careful design and implementation. For example, payments should be risk adjusted to avoid selecting healthy patients so as to properly compensate providers for high-cost patients. Exception reporting, where providers can exclude patients from the denominator of payment calculations, can be avoided by paying only for the numerator, i.e. a payment per patient hitting a target rather than for the proportion of patients hitting a target (Twardella and Brenner 2007). Schemes should also reward measured improvements in quality between two periods rather than the achievement of a given level of quality (Young, Scott, and Best 2010).

Performance-based pay has been promoted extensively to improve health worker performance in LMICs (Loevinsohn and Harding 2005), and its effectiveness has been tested in a range of settings (some examples include Basinga *et al.* 2011; Lundberg, Marek, and Pariyo 2007), with some evidence that it can improve performance. However, there are concerns about the methods applied. Most studies evaluate the implementation of multiple interventions introduced in parallel and cannot distinguish the effects of each (Eichler, Levine, and Performance-Based Incentives Working Group 2009; Galvin 2006). In addition, most evidence derives from rather rudimentary program evaluation, making it “difficult to distinguish proven findings from enthusiasm and marketing” (Galvin 2006, 127S).

### ***The Impact of Incentives within Career Structures on Productivity and Performance***

The field of personnel economics blends the findings of labor economics with human resource management, recognizing that the structure of careers and promotion within organizations can also influence incentives for effort and advancement (Gibbons and Waldman 1999). Incentives in the general labor market are heavily influenced by organizations that focus on their own internal labor market;

in particular, where individual performance in complex jobs is difficult to observe and measure, incentives within career structures become more relevant.

Promotion hierarchies are contained within structures of salary scales and increments, and how they are combined. With subjective performance appraisal methods, these can create incentives for improved performance (Prendergast 1999). Economists view promotion as a competitive process, and so the design of “promotion tournaments” becomes important in providing incentives for worker performance (Lazear and Rosen 1981). This also has implications for team-working in organizations and how incentives within teams operate.

Though salaried payment is often criticized for having no incentives for additional effort or the provision of quality of care, once this is examined in a dynamic context (i.e. promotion next year depends on performance now), it becomes clear that financial incentives can exist to gain promotion and advancement. Incentives for performance are therefore embedded in salary scales for employees (with gaps between each increment in the scale) and in promotion opportunities, creating financial incentives for improved performance and for clear definition of career trajectories.

For some health professionals, these career structures are not well developed and so incentives are not strong for increased performance. Unions prefer equity of pay through “short” pay scales with small gaps between each increment, while employers prefer longer scales with larger gaps and larger wage spreads to encourage higher and increasing levels of performance. Comparing the distribution of wages within firms therefore provides information on the extent to which pay might be linked to performance. There is evidence on these issues in other industries, but very little in health care (Mavromaras and Scott 2005).

Salaried payment and promotion are also prone to bias and “currying favor” with superiors, which may explain the limited use of merit-based promotion and advancement in low-income countries’ health systems, in which transparency of promotion processes is particularly difficult to achieve and the legitimacy of the judgments to be made are thus open to question. There is evidence from such contexts that health workers dislike performance-based reward and prefer seniority (time-served) promotion criteria (for example Ssengooba, McPake, and Palmer 2012).

### **Health Worker Motivation**

*Summary: Health workers are motivated by factors other than money. “Pro-social,” “intrinsic,” and “public service” motivation have all been used to recognize that other, nonpecuniary, aspects of work, jobs, and organizational goals matter in explaining economic behavior.*

*LMICs have been a focus of studies on motivation largely outside the economics literature. For example, health workers in Benin and Kenya reported that alongside predictable constraints, features of organizational culture that penalized them for deviating from norms of working practice also influenced their motivation.*

As seen, health workers are motivated by factors apart from money. Given that wages may not always be flexible or reflect consumer preferences, these other factors will not be captured by wages and are therefore important to examine in their own right. Though this topic departs from traditional labor economics approaches, there is a growing literature in behavioral economics that deals with the more detailed motivations of individual decision making, and borrows many concepts from psychology and neuroscience (Fehr and Falk 2002; Rebitzer and Taylor 2011).

At the heart of many studies of the effects of financial incentives on physician and health care provider behavior lies a utility function that includes not only net income (profit) but also patient's health status or utility (Ellis and McGuire 1990; Evans 1984; Feldstein 1970; McGuire 2000; Siciliani 2009). Physicians' preferences for net income and patients' health status, and how they may be traded off, are a key source of variation in their responses to financial incentives, and were at the heart of debates in the 1970s and 1980s on supplier-induced demand and the principal-agent relationship.

More broadly, there may also be other arguments in the physician's utility function such as intellectual stimulation, reputation, the intrinsic motivation from doing a good job, and other characteristics of the job (Scott 2001). Presumably, physicians with a relatively low marginal utility of net income attach a higher marginal utility to other aspects of work and life, including the intrinsic features of medical practice, altruism, and improving patients' health.

The relative importance of these will influence the response to incentives and therefore the design of remuneration schemes. This has been shown in theoretical models. Financial incentives may not work for doctors with a relatively low marginal utility of income and a high concern for patients' health, or the financial incentive may need to be higher to elicit a response. In particular, if a physician places a relatively heavy weight on improving patients' health, there is a need for much less complex remuneration and payment schedules (Mooney and Ryan 1993).

A parallel literature has also developed within behavioral economics that recognizes both the importance and richness of sources of motivation other than income, and that concepts from psychology can be used to enhance the explanatory power of economic models. Pro-social, intrinsic, and public service motivation have all been used to recognize that other, nonpecuniary, aspects of work, jobs, and organizational goals matter in explaining economic behavior (Fehr and Camerer 2007; Fehr and Falk 2002; Frey 1997; Frey and Jegen 2001).

In addition, in the new field of neuroeconomics, the prefrontal cortex in the brain has been shown to play an important role in resolving conflicts and trade-offs between selfish and pro-social rewards (Fehr and Camerer 2007). This newer literature fits neatly into the older health economics literature about physicians caring not just about income but also about patients' health status.

Though these issues have been recognized in theoretical models in health economics for decades, empirical economics work on measuring these issues is in its infancy. Pro-social motivation has been measured in laboratory experiments,



but only one paper has used medical students in such an experiment to measure the degree of altruism (Godager and Weisen 2011). They examine the marginal rate of substitution between profit and patient health benefit for 42 medical students. Unsurprisingly perhaps, altruism was found to be important and the majority of students placed more weight on altruism than profit, but with substantial heterogeneity in the degree of altruism among all students.

More studies have examined heterogeneity in the monetary motivation of physicians in an indirect way, by using proxy variables. Rizzo and Zeckhauser (2003, 2007), for instance, use a question in a survey that asks doctors what they think their income should be given their career stage, and define this as a reference (or target) income. They show that those whose actual income is below this reference income have stronger growth in income over time, and show that women do not respond to reference incomes.

Iversen and Lurås (2000) use information from the introduction of a capitation scheme in Norway, where general practitioners had to state the preferred number of patients on their list. General practitioners who were allocated less than their preferred number of patients (i.e. a shortage of patients) were found to provide more services to their patients. Whynes, Ennew, and Feighan (1999) examined “entrepreneurial” motives among general practitioners deciding whether to become fund holders (i.e. hold their own devolved budget to pay for hospital services and prescribing) in the United Kingdom. They used a 21-item instrument measuring entrepreneurial traits, and found that these traits influenced that decision.

The growing literature on DCEs is also a source of evidence on the contents of physician utility functions and heterogeneity in their preferences for income. Some of these studies have examined the extent to which the marginal utility of income varies across subgroups of physicians. The marginal utility has been found higher for males (Chomitz *et al.* 1998; Wordsworth *et al.* 2004), for those with less experience (Scott 2001), and for older doctors (Ubach *et al.* 2003).

In LMICs, there have been a number of studies in health worker motivation largely outside the economics literature. Mathauer and Imhoff (2006) report a study using in-depth interviews with nurses and doctors in Benin and Kenya aiming to assess the role of nonfinancial incentives in influencing motivation. Health workers in both countries recognized that they were frequently demotivated and frustrated, and identified the roots of these problems as an inability to meet their professional goals given features of their environment that make that difficult. Alongside predictable constraints such as inadequate supplies and equipment was the poor implementation of human resources management tools, and features of organizational culture that penalized individuals for deviating from norms of working practice. These features are hard to assimilate in a health labor market analysis. They imply that noneconomic variables may dominate in determining the quality of labor supply.

In a mixed-methods study involving interviews, observation, and a survey, Chandler *et al.* (2009) also recognized a range of noneconomic variables playing

important roles in the determination of motivational levels among nonphysician clinicians in two regions of Tanzania, but found that salary considered adequate, and commensurate with social status expectations, was a clear prerequisite for motivation, suggesting in contrast that a labor market framework may be of some use.

Some studies have sought to measure productivity or performance of health staff. Vujicic, Ohiri, and Sparkes (2009), for example, explore the applicability of a productivity measurement tool in Ghana. They find a skewed distribution across districts, indicating that the majority of districts are significantly less productive than is apparently feasible. However, their notion of a “productivity frontier,” which is only implicit in these authors’ argument, does not recognize variable productivity potential at district level; the analysis proves to be highly sensitive to the weighting formula applied to different outputs; and there appear to be data quality issues. The authors conclude that further work is needed to develop a more sophisticated weighting scheme.

Underperformance relative to ideals or norms might be explained by capacity deficits or by motivational deficits. Leonard and Masatu (2010) explore the distinction by exploiting what they call the Hawthorne effect, in interpretation, the difference between the performance of staff when they know they are observed and when they do not, to distinguish between the two. For doctors operating in the Arusha region of Tanzania, they find a variable degree of difference between the two sets of observations (and a regression to prior performance levels over time indicating that the effect of observation diminishes over time), which is smaller for nonpublic facilities than public facilities and smaller in facilities with greater decentralization of decision making on, for example, hiring and firing and setting salary levels.

These findings are consistent with those of Das and Hammer (2004, 2005) who use theoretical knowledge tested by vignettes as their basis, and find significant performance problems in Delhi, India, correlated with provider type and location with facilities in richer areas and the private sector containing higher performing doctors. These findings contrast with those of Das and Sohnesen (2007) in Paraguay, however, where Ministry of Health facilities performed better in relation to a doctor effort index than facilities associated with Paraguay’s social security system.

## Note

1. There is a large literature in labor economics concerning the impact of trade unions on wages and employment (Ehrenberg and Schwarz 1986; Lewis 1986).

## References

- Acemoglu, D., and D. Autor. 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 1043–171. Elsevier.



- Akwataghibe, N., U. Chiroma, A. Ogbeide, E. Agbo, I. Aladeniyi, and A. Vincent. 2012. "Assessing Human Resources for Health Revenues and Accountability in Nigeria." Royal Tropical Institute, KIT Development Policy and Practice, Amsterdam.
- Antonazzo, E., A. Scott, D. Skatun, and R. F. Elliott. 2003. "The Labour Market for Nursing: A Review of the Labour Supply Literature." *Health Economics* 12: 465–78.
- Antwi, J., and D. C. Phillips. 2013. "Wages and Health Worker Retention: Evidence from Public Sector Wage Reforms in Ghana." *Journal of Development Economics* 102: 101–15.
- Appiah-Denkyira, E., C. H. Herbst, A. Soucat, C. Lemiere, and K. Saleh. 2013. *Towards Interventions on Human Resources for Health in Ghana: Evidence for Health Workforce Planning and Results*. Washington, DC: World Bank.
- Attah, R., T. Lievens, M. Vujicic, and J. Brown-Annan. 2010. "Health Worker Attitudes toward Rural Service in Liberia: Results from Qualitative Research." Health, Nutrition and Population (HNP) Discussion Paper, World Bank, Washington, DC.
- Bach, S. 2007. "Going Global? The Regulation of Nurse Migration in the UK." *British Journal of Industrial Relations* 45: 383–403.
- Baltagi, B. H., E. Bratberg, and T. H. Holmås. 2005. "A Panel Data Study of Physicians' Labor Supply: The Case of Norway." *Health Economics* 14: 1035–45.
- Barnighausen, T., and D. Bloom. 2009. "Financial Incentives for Return of Service in Underserved Areas: A Systematic Review." *BMC Health Services Research* 9: 86.
- Barr, A., M. Lindelow, and P. Serneels. 2009. "Corruption in Public Service Delivery: An Experimental Analysis." *Journal of Economic Behavior and Organization* 72: 225–39.
- Basinga, P., P. J. Gertler, A. Binagwaho, A. L. B. Soucat, J. Sturdy, and C. M. J. Vermeersch. 2011. "Effect on Maternal and Child Health Services in Rwanda of Payment to Primary Health-Care Providers for Performance: An Impact Evaluation." *The Lancet* 377: 1421–28.
- Birrell, R. J. 2004. "Australian Policy on Overseas-Trained Doctors." *Medical Journal of Australia* 181: 635–39.
- Blau, F. D., and L. M. Kahn. 1999. "Institutions and Laws in the Labor Market." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 1399–461. Elsevier.
- Blundell, R., and T. Macurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 1559–695. Elsevier.
- Buchan, J., B. McPake, K. Mensah, and G. Rae. 2009. "Does a Code Make a Difference—Assessing the English Code of Practice on International Recruitment." *Human Resources for Health* 7: 33.
- Burstein, P. L., and J. Cromwell. 1985. "Relative Incomes and Rates of Return for U.S. Physicians." *Journal of Health Economics* 4: 63–78.
- Buykx, P., J. Humphreys, J. Wakerman, and D. Pashen. 2010. "Systematic Review of Effective Retention Incentives for Health Workers in Rural and Remote Areas: Towards Evidence-Based Policy." *The Australian Journal of Rural Health* 18: 102–09.
- Campbell, S., A. Scott, R. Parker, L. Naccarella, J. Furler, D. Young, and P.M. Sivey. 2010. "Implementing Pay-For-Performance in Australian Primary Care: Lessons from the United Kingdom and the United States." *Medical Journal of Australia* 193: 408–11.
- Card, D. 1999. "The Causal Effect of Education on Earnings." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 1801–63. Elsevier.

- Carnoy, M., P. Loyalka, G. V. Androushchak, and A. Proudnikova. 2012. "The Economic Returns to Higher Education in the BRIC Countries and Their Implications for Higher Education Expansion." CEPA Working Paper Series, Center for Education Policy Analysis, Stanford University.
- Celletti, F., T. A. Reynolds, A. Wright, A. Stoertz, and M. Dayrit. 2011. "Educating a New Generation of Doctors to Improve the Health of Populations in Low- and Middle-Income Countries." *PLoS Med* 8.
- Chandler, C. I., S. Chonya, F. Mtei, H. Reyburn, and C. J. Whitty. 2009. "Motivation, Money and Respect: A Mixed-Method Study of Tanzanian Non-Physician Clinicians." *Social Science & Medicine* 68: 2078–88.
- Cheng, T. C., A. Scott, S.-H. Jeon, G. Kalb, J. Humphreys, and C. Joyce. 2012. "What Factors Influence the Earnings of General Practitioners and Medical Specialists? Evidence from the Medicine in Australia: Balancing Employment and Life Survey." *Health Economics* 21: 1300–17.
- Chilopora, G., C. Pereira, F. Kamwendo, A. Chimbiri, E. Malunga, and S. Bergstrom. 2007. "Postoperative Outcome of Caesarean Sections and Other Major Emergency Obstetric Surgery by Clinical Officers and Medical Officers in Malawi." *Human Resources for Health* 5: 17.
- Chomitz, K., G. Setiadi, A. Azwar, M. Ismail, and P. Widiyarti. 1998. "What Do Doctors Want? Developing Incentives for Doctors to Serve in Indonesia's Rural and Remote Areas." Policy Research Working Paper 1888, World Bank, Washington, DC.
- Cunich, M., and S. Whelan. 2010. "Nurse Education and the Retention of Registered Nurses in New South Wales." *Economic Record* 86: 396–413.
- Das, J., and J. Hammer. 2004. "Which Doctor? Combining Vignettes and Item-Response to Measure Doctor Quality." Policy Research Working Paper 3301, World Bank, Washington, DC.
- . 2005. "Money for Nothing: The Dire Straits of Medical Practice in Delhi, India." Policy Research Working Paper 3669, World Bank, Washington, DC.
- Das, J., and T. P. Sohnesen. 2007. "Variations in Doctor Effort: Evidence from Paraguay." *Health Affairs (Millwood)* 26: w324–37.
- Department of Health. 2001. *Eccles Code of Practice for NHS Employers Involved in the International Recruitment of Healthcare Professionals*. London.
- . 2004. *Code of Practice for the International Recruitment of Healthcare Professionals*. London.
- Eccles, M., S. Shepperd, A. Scott, G. Flodgren, E. Parmelli, and F. Beyer. 2010. "An Overview of Reviews Evaluating the Effects of Financial Incentives in Changing Healthcare Professional Behaviours and Patient Outcomes (Protocol)." *Cochrane Database of Systematic Reviews* 7.
- Eggleston, K., and A. Bir. 2006. "Physician Dual Practice." *Health Policy* 78: 157–66.
- Ehrenberg, R. G., and J. L. Schwarz. 1986. "Public-Sector Labor Markets." In *Handbook of Labor Economics*, edited by C. A. Orley and L. Richard, 1219–60. Elsevier.
- Eichler, R., R. Levine, and Performance-Based Incentives Working Group, ed. 2009. *Performance Incentives for Global Health: Potential and Pitfalls*. Washington, DC: Center for Global Development.
- Elliott, R. F., A. H. Y. Ma, A. Scott, D. Bell, and E. Roberts. 2007. "Geographically Differentiated Pay in the Labour Market for Nurses." *Journal of Health Economics* 26: 190–212.

- Ellis, R., and T. McGuire. 1990. "Optimal Payment Systems for Health Services." *Journal of Health Economics* 9: 375–96.
- Evans, R. 1984. *Strained Mercy: The Economics of Canadian Health Care*. Toronto: Butterworths.
- Fehr, E., and C. Camerer. 2007. "Social Neuroeconomics: The Neural Circuitry of Social Preferences." *Trends in Cognitive Sciences* 11: 419–27.
- Fehr, E., and A. Falk. 2002. "Psychological Foundations for Incentives." *European Economic Review* 46: 687–724.
- Feldstein, M. 1970. "The Rising Price of Physicians' Services." *Review of Economics and Statistics* (May) 52: 121–33.
- Ferrinho, P., M. Sidat, M. Fresta, A. Rodrigues, I. Fronteira, F. da Silva, H. Mercer, J. Cabral, and G. Dussault. 2011. "The Training and Professional Expectations of Medical Students in Angola, Guinea-Bissau and Mozambique." *Human Resources for Health* 9: 9.
- Frey, B. 1997. *Not Just for the Money: An Economic Theory of Personal Motivation*. Edward Elgar Publishing.
- Frey, B., and R. Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys* 15: 589–611.
- Gaal, P., T. Evetovits, and M. McKee. 2006. "Informal Payment for Health Care: Evidence from Hungary." *Health Policy* 77: 86–102.
- Galvin, R. S. 2006. "Evaluating the Performance of Pay for Performance." *Medical Care Research and Review* 63: 126S–30.
- Garcia-Prado, A., and P. Gonzalez. 2007. "Policy and Regulatory Responses to Dual Practice in the Health Sector." *Health Policy* 84: 142–52.
- Gessessew, A., G. A. Barnabas, N. Prata, and K. Weidert. 2011. "Task Shifting and Sharing in Tigray, Ethiopia, to Achieve Comprehensive Emergency Obstetric Care." *International Journal of Gynaecology and Obstetrics: The Official Organ of the International Federation of Gynaecology and Obstetrics* 113: 28–31.
- Gibbons, R., and M. Waldman. 1999. "Careers in Organizations: Theory and Evidence." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 2373–437. Elsevier.
- Godager, G., and D. Weisen. 2011. *Profit or Patients' Health Benefit? Exploring the Heterogeneity in Physician Altruism*. Oslo: Health Economics Research Program, University of Oslo.
- González, P., and I. Macho-Stadler. 2013. "A Theoretical Approach to Dual Practice Regulations in the Health Sector." *Journal of Health Economics* 32: 66–87.
- Gosden T., F. Forland, I. S. Kristiansen, M. Sutton, B. Leese, A. Giuffrida, M. Sergison, and L. Pedersen. 2000. "Capitation, Salary, Fee-For-Service and Mixed Systems of Payment: Effects on the Behaviour of Primary Care Physicians." *Cochrane Database of Systematic Reviews*.
- . 2001. "Impact of Payment Method on Behaviour of Primary Care Physicians: A Systematic Review." *Journal of Health Services Research and Policy* 6: 44–55.
- Grobler, L., B. Marais, S. Mabunda, P. Marindi, H. Reuter, and J. Volmink. 2009. "Interventions for Increasing the Proportion of Health Professionals Practising in Rural and Other Underserved Areas." *Cochrane Database of Systematic Reviews* 1: CD005314. doi: 10.1002/14651858.CD005314.pub2.

- Gruen, R., R. Anwar, T. Begum, J. R. Killingsworth, and C. Normand. 2002. "Dual Job Holding Practitioners in Bangladesh: An Exploration." *Social Science & Medicine* 54: 267–79.
- Hanel, B., G. Kalb, and A. Scott. 2012. "Nurses' Labour Supply Elasticities: The Importance of Accounting for Extensive Margins." Working Paper 9, The University of Melbourne, Melbourne.
- Hanson, K., and W. Jack. 2010. "Designing Incentives for Rural Health Workers." In *Incentives and Dynamics in the Ethiopian Health Worker Labor Market*, edited by W. Jack, J. De Laat, K. Hanson, and A. Soucat, Chapter 3. Washington, DC: World Bank.
- Herbst, C. H., M. Vledder, K. Campbell, M. Sjöblom, and A. Soucat. 2011. *The Human Resources for Health Crisis in Zambia: An Outcome of Health Worker Entry, Exit, and Performance within the National Labor Health Market*. Africa Human Development Series. Washington, DC: World Bank.
- Hicks, V., and O. Adams. 2001. "Pay and Non-Pay Incentives, Performance and Motivation." Prepared for the Global Health Workforce Strategy Group, WHO, Geneva.
- Holmstrom, B., and P. Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics and Organisation* 7: 24–52.
- Hongoro, C., and C. Normand. 2006. "Health Workers: Building and Motivating the Workforce." In *Disease Control Priorities in Developing Countries*. 2nd ed., edited by D. T. Jamison, J. G. Breman, A. R. Measham, G. Alleyne, M. Claeson, D. B. Evans, P. Jha, A. Mills, and P. Musgrove. New York: Oxford University Press.
- Hounton, S., D. Newlands, N. Meda, and V. De Brouwere. 2009. "A Cost-Effectiveness Study of Caesarean-Section Deliveries by Clinical Officers, General Practitioners and Obstetricians in Burkina Faso." *Human Resources for Health* 7: 34.
- Huicho, L., R. W. Scherpbier, A. M. Nkowane, and C. G. Victora. 2008. "How Much Does Quality of Child Care Vary Between Health Workers with Differing Durations of Training? An Observational Multicountry Study." *The Lancet* 372: 910–16.
- Hunt, J. 2010. "Bribery in Health Care in Uganda." *Journal of Health Economics* 29: 699–707.
- Iversen, T., and H. Lurås. 2000. "Economic Motives and Professional Norms: The Case of General Medical Practice." *Journal of Economic Behavior & Organization* 43: 447–70.
- Jumpa, M., S. Jan, and A. Mills. 2007. "The Role of Regulation in Influencing Income-Generating Activities among Public Sector Doctors in Peru." *Human Resources for Health* 5: 5.
- Kernick, D., and A. Scott. 2002. "Economic Approaches to Doctor/Nurse Skill Mix: Problems, Pitfalls, and Partial Solutions." *British Journal of General Practice* 52: 42–46.
- Kruk, M. E., J. C. Johnson, M. Gyakobo, P. Agyei-Baffour, K. Asabir, S. R. Kotha, J. Kwansah, E. Nakua, R. C. Snow, and M. Dzodzomenyo. 2010. "Rural Practice Preferences among Medical Students in Ghana: A Discrete Choice Experiment." *Bulletin of the World Health Organization* 88: 333–41.
- Kurowski, C., C. Carpio, M. Vujicic, L. O. Gostin, and T. Baytor. 2012. *Towards a Regional Strategy to Strengthen the Nurse Workforce of the English-Speaking Caricom: International Legal Instruments, Agreements and Obligations*. Washington, DC: World Bank.
- Lagarde, M., D. Blaauw, and J. Cairns. 2012. "Cost-Effectiveness Analysis of Human Resources Policy Interventions to Address the Shortage of Nurses in Rural South Africa." *Social Science & Medicine* 75: 801–06.

- Langwell, K. M. 1982. "Differences by Sex in Economic Returns Associated with Physician Specialization." *Journal of Health Politics, Policy and Law* 6: 752–61.
- Laurant, M., D. Reeves, R. Hermens, J. Braspenning, R. Grol, and B. Sibbald. 2004. "Substitution of Doctors by Nurses in Primary Care." *Cochrane Database of Systematic Reviews* 4.
- Lazear, E. P., and S. Rosen. 1981. "Rank-Order Tournaments as Optimum Labor Contracts." *Journal of Political Economy* 89: 841–64.
- Lemiere, C., C. H. Herbst, N. Jahanshahi, E. Smith, and A. Soucat. 2011. *Reducing Geographical Imbalances of Health Workers in Sub-Saharan Africa: A Labor Market Perspective on What Works, What Does Not, and Why*. Africa Human Development Series. Washington, DC: World Bank.
- Leonard, K. L., and M. C. Masatu. 2010. "Using the Hawthorne Effect to Examine the Gap between a Doctor's Best Possible Practice and Actual Performance." *Journal of Development Economics* 93: 226–34.
- Lewis, G. 1986. "Union Relative Wage Effects." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and R. Layard, Chapter 20, 1139–81. Elsevier.
- Loevinsohn, B., and A. Harding. 2005. "Buying Results? Contracting for Health Service Delivery in Developing Countries." *The Lancet* 366: 676–81.
- Lori, J. R., S. Rominski, J. Richardson, P. Agyei-Baffour, N. E. Kweku, and M. Gyakobo. 2012. "Factors Influencing Ghanaian Midwifery Students' Willingness to Work in Rural Areas: A Computerized Survey." *International Journal of Nursing Studies* 49: 834–41.
- Lundberg, M., T. Marek, and G. Pariyo. 2007. "Contracting for Primary Health Care in Uganda." Health Systems for Outcomes Publication, World Bank, Washington, DC.
- Makasa, E. 2009. "The Human Resource Crisis in the Zambian Health Sector—A Discussion Paper." *Medical Journal of Zambia* 35: 81–87.
- Malcomson, M. 1999. "Individual Employment Contracts." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 2291–372. Elsevier.
- Mangham, L. J., and K. Hanson. 2008. "Employment Preferences of Public Sector Nurses in Malawi: Results from a Discrete Choice Experiment." *Tropical Medicine and International Health* 13: 1433–41.
- Masango, S., K. Gathu, and S. Sibandze. 2008. *Retention Strategies for Swaziland's Health Sector Workforce: Assessing the Impact of Non-Financial Incentives*. EQUINET Discussion Paper Series, Harare: EQUINET, ECSA-HC.
- Mathauer, I., and I. Imhoff. 2006. "Health Worker Motivation in Africa: The Role of Non-Financial Incentives and Human Resource Management Tools." *Human Resources for Health* 4: 24.
- Mavromaras, K., and A. Scott. 2005. "Promotion to Hospital Consultant in NHS Scotland." *International Journal of Manpower* 26: 660–72.
- McCoy, D., S. Bennett, S. Witter, B. Pond, B. Baker, J. Gow, J. G. S. Chand, T. Ensor, and B. McPake. 2008. "Salaries and Incomes of Health Workers in Sub-Saharan Africa." *The Lancet* 371: 675–81.
- McGuire, T. 2000. "Physician Agency." In *Handbook of Health Economics*, edited by J. Newhouse and A. J. Culyer. Amsterdam: Elsevier.
- Merrell, K., and R. Berenson. 2010. "Structuring Payment for Medical Homes." *Health Affairs* 29: 5.



- Mooney, G., and M. Ryan. 1993. "Agency in Health Care: Getting Beyond First Principles." *Journal of Health Economics* 12: 125.
- Moretti, E. 2011. Local Labor Markets. In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 1237–313. Elsevier.
- Munga, M. A., and D. R. Mbilinyi. 2008. *Non-Financial Incentives and Retention of Health Workers in Tanzania: Combined Evidence from Literature Review and a Focussed Crosssectional Study*. EQUINET Discussion Paper Series. Harare: NIMRI, ECSA-HC, EQUINET.
- Ndeti, D. M., L. Khasakhala, and J. O. Omolo. 2008. *Incentives for Health Worker Retention in Kenya: An Assessment of Current Practice*. EQUINET Discussion Paper Series. EQUINET with African Mental Health Foundation, University of Namibia, Training and Research Support Centre, University of Limpopo and Harare: ECSA-Regional Health Community, EQUINET.
- Nicholson, S. 2002. "Physician Specialty Choice under Uncertainty." *Journal of Labor Economics* 20: 816–47.
- Nicholson, S., and C. Propper. 2011. "Medical Workforce." In *Handbook of Health Economics*, edited by M. Pauly, T. McGuire, and P. Barros, Chapter 14, 873–925. Elsevier.
- Nickell, S., and R. Layard. 1999. "Labor Market Institutions and Economic Performance." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 3029–84. Elsevier.
- OECD (Organisation for Economic Co-operation and Development). 2002. *Trends in International Migration: SOPEMI 2002*. Paris: OECD.
- Onwujekwe, O., N. Dike, B. Uzochukwu, and O. Ezeoke. 2010. "Informal Payments for Healthcare: Differences in Expenditures from Consumers and Providers Perspectives for Treatment of Malaria in Nigeria." *Health Policy* 96: 72–79.
- Onzubo, P. 2007. "Turnover of Health Professionals in the General Hospitals in West Nile Region." *Health Policy and Development* 5: 28–34.
- Penn-Kekana, L., D. Blaauw, K. S. Tint, D. Monareng, and J. Chege. 2005. *Nursing Staff Dynamics and Implications for Maternal Health Provision in Public Health Facilities in the Context of HIV/AIDS*. Frontiers in Reproductive Health, Population Council.
- Pond, B., and B. McPake. 2006. "The Health Migration Crisis: The Role of Four Organisation for Economic Co-operation and Development Countries." *The Lancet* 367: 1448–55.
- Prendergast, C. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37: 7–63.
- Propper, C., and J. Van Reenen. 2010. "Can Pay Regulation Kill? Panel Data Evidence on the Effect of Labor Markets on Hospital Performance." *Journal of Political Economy* 118: 222–73.
- Qin, X., L. Li, and C.-R. Hsieh. 2013. "Too Few Doctors or Too Low Wages? Labor Supply of Health Care Professionals in China." *China Economic Review* 24: 150–64.
- Raha, S., P. Berman, and A. Bhatnagar. 2009a. *Some Priority Challenges of the Nursing Sector in India*. Washington, DC: World Bank.
- . 2009b. *Career Preferences of Medical and Nursing Students in Uttar Pradesh*. Washington, DC: World Bank.
- Rao, K. D., S. Ramani, S. Murthy, I. Hazarika, N. Khandpur, M. Chokshi, S. Khana, M. Vujicic, P. Berman, and M. Ryan. 2010. *Health Worker Attitudes toward Rural Service in India: Results from Qualitative Research*. Washington, DC: World Bank.

- Rebitzer, J. B., and L. J. Taylor. 2011. "Extrinsic Rewards and Intrinsic Motives: Standard and Behavioral Approaches to Agency and Labor Markets." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, Chapter 8, 701–72. Elsevier.
- Rizzo, J. A., and R. J. Zeckhauser. 2003. "Reference Incomes, Loss Aversion, and Physician Behavior." *Review of Economics and Statistics* 85: 909–22.
- . 2007. "Pushing Incomes to Reference Points: Why Do Male Doctors Earn More?" *Journal of Economic Behavior and Organization* 63: 514–36.
- Robinson, J. 2001. "Theory and Practice in the Design of Physician Payment Incentives." *Milbank Quarterly* 79 (2): 149–77.
- Rokx, C., J. Giles, E. Satriawan, P. Marzoeki, P. Harimurti, and E. Yavuz. 2010. *New Insights into the Provision of Health Services in Indonesia: A Health Workforce Study*. Washington, DC: World Bank.
- Roland, M. 2004. "Linking Physicians' Pay to the Quality of Care—A Major Experiment in the United Kingdom." *The New England Journal of Medicine* 351: 1448–54.
- Rosen, S. 1986. "The Theory of Equalizing Differences." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and R. Layard, 641–92. Elsevier.
- Rouleau, D., P. Fournier, A. Philibert, B. Mbengue, and A. Dumont. 2012. "The Effects of Midwives' Job Satisfaction on Burnout, Intention to Quit and Turnover: A Longitudinal Study in Senegal." *Human Resources for Health* 10: 9.
- Runnels, V., R. Labonte, and C. Packer. 2011. "Reflections on the Ethics of Recruiting Foreign-Trained Human Resources for Health." *Human Resources for Health* 9: 2.
- Russo, G., B. McPake, I. Fronteira, and P. Ferrinho. 2013. "Negotiating Markets for Health: An Exploration of Physicians' Engagement in Dual Practice in Three African Capital Cities." *Health Policy and Planning*, September 26: 1–10. <http://heapol.oxfordjournals.org/content/early/2013/09/26/heapol.czt071.short>.
- Sæther, E. 2005. "Physicians' Labour Supply: The Wage Impact on Hours and Practice Combinations." *Labour* 19: 673–703.
- Savage, E., K. van Gool, M. Haas, R. Viney, and M. Vu. 2009. *Extended Medicare Safety Net: Review Report*. Canberra: Australian Government Department of Health and Ageing.
- Scott, A. 2001. "Eliciting GPs' Preferences for Pecuniary and Non-pecuniary Job Characteristics." *Journal of Health Economics* 20: 329–47.
- . 2006. "The Productivity of the Health Workforce." *Australian Economic Review* 39: 312–17.
- Scott A, P. Sivey, D. Ait Ouakrim, L. Willenberg, L. Naccarella, J. Furler, and D. Young. 2011. "The Effect of Financial Incentives on the Quality of Health Care Provided by Primary Care Physicians." *Cochrane Database of Systematic Reviews* 4.
- Serra, D., P. Serneels, M. Lindelow, and J. G. Montalvo. 2010. *Discovering the Real World: Health Workers' Career Choices and Early Work Experience in Ethiopia*. Africa Human Development Series. Washington, DC: World Bank.
- Shields, M. A. 2004. "Addressing Nurse Shortages: What Can Policy Makers Learn from the Econometric Evidence on Nurse Labour Supply?" *Economic Journal* 114 (499): F464–98.
- Siciliani, L. 2009. "Paying for Performance and Motivation Crowding Out." *Economics Letters* 103: 68–71.
- Sivey, P., A. Scott, J. Witt, C. Joyce, and J. Humphreys. 2012. "Junior Doctors' Preferences for Specialty Choice." *Journal of Health Economics* 31: 813–23.

- Socha, K. Z., and M. Bech. 2011. "Physician Dual Practice: A Review of Literature." *Health Policy* 102: 1–7.
- Spetz, J. 2002. "The Value of Education in a Licensed Profession: The Choice of Associate or Baccalaureate Degrees in Nursing." *Economics of Education Review* 21: 73–85.
- Ssengooba, F., B. McPake, and N. Palmer. 2012. "Why Performance-Based Contracting Failed in Uganda—An "Open-Box" Evaluation of a Complex Health System Intervention." *Social Science & Medicine* 75: 377–83.
- Sutton, M., R. Elder, B. Guthrie, and G. Watt. 2009. "Record Rewards: The Effects of Targeted Quality Incentives on the Recording of Risk Factors by Primary Care Providers." *Health Economics* 19(1):1–13.
- Twardella, D., and H. Brenner. 2007. "Effects of Practitioner Education, Practitioner Payment and Reimbursement of Patients' Drug Costs on Smoking Cessation in Primary Care: A Cluster Randomised Trial." *Tobacco Control* 16: 15–21.
- Ubach, C., A. Scott, F. French, M. Awramenko, and G. Needham. 2003. "What Do Hospital Consultants Value About Their Jobs? A Discrete Choice Experiment." *British Medical Journal* 326: 1432.
- Vujicic, M., M. Alfano, and B. Shengelia. 2010. "Getting Health Workers to Rural Areas: Innovative Analytic Work to Inform Policy Making." Health, Nutrition and Population (HNP) Discussion Paper, World Bank, Washington, DC.
- Vujicic, M., K. Ohiri, and S. Sparkes. 2009. *Working in Health: Financing and Managing the Public Sector Health Workforce*. World Bank: Washington, DC.
- Vujicic, M., B. Shengelia, M. Alfano, and H. B. Thu. 2011. "Physician Shortages in Rural Vietnam: Using a Labor Market Approach to Inform Policy." *Social Science & Medicine* 73: 970–77.
- Vujicic, M., S. Sparkes, and S. Mollahaliloglu. 2009. *Health Workforce Policy in Turkey: Recent Reforms and Issues for the Future*. Washington, DC: World Bank.
- Vujicic, M., P. Zurn, K. Diallo, O. Adams, and M. R. Dal Poz. 2004. "The Role of Wages in the Migration of Health Care Professionals from Developing Countries." *Human Resources for Health* 2: 3.
- Warriner, I., O. Meirik, M. Hoffman, C. Morroni, J. Harries, N. My Huong, Vy N.D., Seuc A.H. 2006. "Rates of Complication in First-Trimester Manual Vacuum Aspiration Abortion Done by Doctors and Mid-Level Providers in South Africa and Vietnam: A Randomised Controlled Equivalence Trial." *The Lancet* 368: 1965–72.
- Whalley, D., H. Gravelle, and B. Sibbald. 2008. "Effect of the New Contract on GPs' Working Lives and Perceptions of Quality of Care: A Longitudinal Survey." *British Journal of General Practice* 58: 8–14.
- WHO (World Health Organization). 2006. *The World Health Report: Working Together for Health*. Geneva: WHO.
- . 2010. *WHO Global Code of Practice on the International Recruitment of Health Personnel*. Sixty-third World Health Assembly—WHA63.16. Geneva, Switzerland.
- Whynes, D. K., C. T. Ennew, and T. Feighan. 1999. "Entrepreneurial Attitudes of Primary Health Care Physicians in the United Kingdom." *Journal of Economic Behavior and Organization* 38: 331–47.
- Witter, S., B. Thi Thu Ha, B. Shengalia, and M. Vujicic. 2011. "Understanding the 'Four Directions of Travel': Qualitative Research into the Factors Affecting Recruitment and Retention of Doctors in Rural Vietnam." *Human Resources for Health* 9: 20.



- Wordsworth, S., D. Skåtun, A. Scott, and F. French. 2004. "Preferences for General Practice Jobs: A Survey of Principals and Sessional GPs." *British Journal of General Practice* 54: 740–46.
- World Bank. 2008. *Human Resources for Health in Ethiopia: Summary Report*. Washington, DC: World Bank.
- . 2009. *The Nurse Labor and Education Markets in the English-Speaking Caricom: Issues and Options for Reform*. Washington, DC: World Bank.
- Young, D., A. Scott, and J. D. Best. 2010. "For Love or Money? Changing the Way GPs Are Paid to Provide Diabetes Care." *Medical Journal of Australia* 193: 67–68.



# Analytical Approaches

*Two main types of labor market analysis use an explicit economics framework: descriptive and causal analysis.*

*Descriptive labor market analysis helps establish the nature and extent of labor market disequilibrium, which in turn help inform the identification of key issues, research questions, and further analysis and policies to improve the efficient operation of the labor market. It might provide data about demand, compensation, market structure, supply, interaction of demand and supply, trends, and distribution. There are few descriptive studies of the international labor market that shapes health professional migration trends, but there are more studies at the national level.*

*Causal labor market analysis attempts to identify the effects of changes in labor market supply and demand conditions or regulation on the behavior of both employers and workers, and on the value of these changes to society. Understanding the range of factors that can influence labor market behaviors is essential in developing policies to reduce labor market disequilibrium to improve societal outcomes. Causality can be established using experimental approaches (e.g. randomized control trials), quasi-experimental approaches (e.g. difference-in-difference analysis), and non-experimental approaches (e.g. instrumental variable analysis).*

*Relevant descriptive datasets include the number of health workers of different types, age, gender, hours worked, specialty, and geographic location. Data to examine the causal effects of policy change include these descriptive data, but often contain more variables (and are longitudinal) to enable the researcher to control for as many factors influencing behavior as possible.*

The previous discussion of issues and empirical evidence involved a wide range of analytical approaches used by labor economists that this report cannot discuss in full. However, the application of labor and health economics to the analysis of health care labor markets requires an understanding of the broad types of empirical methods used, and of the types of data that are essential in supporting these analyses. The aim of this section is therefore to broadly summarize the nature and types of analytical approaches that can be used and types of data that are required, and so is more relevant to those wishing to pursue such analyses.

Ideally, it is appealing to develop a theoretical model and use this to generate predictions and hypotheses that can be tested using data and various empirical approaches. Theoretical models examine the conceptual relationships between different “actors” in the market, and model the effect of changes in incentives and other policy drivers on behavior. They often understand situations by simplifying them using a range of assumptions about what is in the utility functions of actors (e.g. consumption, leisure, and patient health status), what time and money constraints they face when maximizing their utility, and how market relationships work. Theory models can be used to make specific predictions that are tested using empirical data that match the variables examined in the theoretical models—these are known as structural models.

The most widely used is the economic theory of labor supply that examines the effect of changes in wages on hours worked (Blundell and Macurdy 1999). Theoretical models may also be developed to analyze a specific problem—we have already discussed a theoretical model of dual practice (González and Macho-Stadler 2013). Testing these models is often difficult as empirical data often do not match what is in the theory model, and so “reduced form” empirical approaches are used. More often, theoretical models provide a broad framework that identifies the key variables, relationships, and assumptions that should be examined in empirical work. In reduced form empirical models, empirical results are more difficult to interpret though much can still be learned through the use of micro-econometric techniques that help to move from measuring associations to identifying causality.

In this section, we focus largely on empirical rather than theoretical and structural approaches. Nevertheless, it is important to be able to use economic thinking and frameworks discussed in previous sections to help guide the design of empirical work and data collection.

As there are a wide range of analytical approaches that depend on the problem and research question at hand (the *Handbook of Labor Economics*, for example, now runs to four volumes), this section summarizes the broad approaches used in labor market analysis. For those who wish to explore these approaches in more detail, see Angrist and Krueger (1999); Blundell and Macurdy (1999); DiNardo and Lee (2011); List and Rasul (2011); and Moffitt (1999).

The type of empirical labor market analysis and data collected should be driven by the policy issues and research questions being addressed. There are two main types of labor market analysis that use an explicit economics framework (Angrist and Krueger 1999): descriptive and causal.

### **Descriptive Labor Market Analysis: The Nature and Extent of Disequilibrium**

Descriptive labor market analysis uses a range of data sources to establish the current state of the labor market or labor market problem: it can therefore be general with the aim of identifying trends and issues, or can focus on an already

known issue and begin to describe its key features and characteristics. Research questions and hypotheses are usually generated through such an analysis of the labor market.

Descriptive analysis helps establish the nature and extent of labor market disequilibrium, which in turn helps in identifying key issues, research questions, further analysis, and, ultimately, policies to improve the efficient operation of the labor market. Disequilibrium, as we have seen, may exist because of the way the health system has been structured and funded, the types (or lack) of institutions and regulations that govern health worker labor markets, and any changes in demand and supply conditions that have altered market forces and the behavior of employers and health workers. These include more general changes in the economy and in government spending that influence the demand for health workers, and so it is often important to look beyond the health care sector. Analyses may consist of some simple cross-tabulations and associations, e.g. comparing wages (and other characteristics) in the public and private sectors, or between urban and rural areas, and how the wage gap is changing over time, and can provide some testable hypotheses that can be explored further with econometric analysis.

It is often hard to identify disequilibrium in the literature, and so a range of data on trends in workforce supply and demand or need is used. Judgments are then frequently made about the existence of a shortage or surplus and how these may develop into the future. In simple terms, one needs to observe trends and changes in wages rates (prices) and employment (quantities) to be able to identify shortage or surplus.

Though many health workforce planning exercises attempt to project demand and supply into the future, these are largely devoid of economic thinking and so ignore important economic information, such as budget constraints and changes in wages or prices and specific measures of shortage, such as vacancies and waiting times (Scott *et al.* 2011). The importance of examining variations or changes in wages can reveal much about the state of the labor market, yet it is surprising that very little is known or collected about wages in most national health workforce data collections. This should be a central item for data collection in any labor market analysis.

Ideally, a descriptive analysis of the whole labor market would seek to cover the range of issues already discussed in the previous sections and summarized in the questions below:

- *Demand.* What are recent trends in population or regional need for health care? How is the health system funded, and how is the health workforce funded and paid for? What are the issues surrounding potential changes in skill mix?
- *Wages, compensation, prices, waiting times, and vacancies.* What are the key indicators of shortage or surplus in the labor market, including differences and movements in wages or prices for services, vacancies, and employment?

Where wages are inflexible, what are the trends in waiting times for specific health workers?

- *Market structure.* How many separate employers are there in the market, and are health workers organized into unions? What is the extent of competition for workers between employers?
- *Supply.* What are the trends in the numbers, characteristics, and distribution of health workers? This is often the main and only information available to governments where health workers are licensed or accredited, but without other data, it is impossible to assess the nature or extent of labor market disequilibriums. Examination of changes in the age and gender composition of health workers may, though, be revealing in terms of identifying shortages, for example, where the workforce is getting older on average, suggesting lower numbers of entrants into the workforce. Other assumptions of the perfectly competitive model are relevant to describe here, such as the extent to which entry into the labor market is restricted through training, accreditation, and licensing.
- *The interaction of demand and supply.* Taking into account information on wages, prices, waiting times, and data on the number of health workers (price and quantity data), is the labor market in surplus or shortage? A shortage would be indicated by rising wages, vacancies, and waiting times, combined with reductions in supply such as falling hours worked or falling numbers of health workers. A surplus would be indicated by falling wages, vacancies, and waiting times, and unemployment of health workers. Given problems of market imperfections and market failure, wages may not be flexible and so not change very much over time. Then the data on waiting times and prices or vacancies may be more reliable indicators of labor market conditions. If wages are not changing over time in response to changes in demand and supply conditions, this is evidence of market rigidities and imperfections that may need to be examined.
- *Trends in the broader labor market.* What is happening in the general labor market, and are these trends influencing the labor market for health workers? A range of other occupations competes for health workers, and so examining broader labor markets trends will also inform what is occurring in health worker labor markets. For example, if wage rates in other occupations are falling relative to those for health workers, this increases the attractiveness of health care as an occupation.
- *Distribution.* How does each of the above vary across submarkets, including geographic markets, public or private markets, specialties, and acute care versus community settings? Wage differentials and movements of workers between these markets may also be indicative of changes in demand and supply conditions.

Below are some examples from low- and middle-income countries (LMICs) of studies that have attempted to provide a “whole of labor market” analysis by answering a range of the above questions. These studies have explicitly taken a labor market approach and have, to some extent, considered the interaction of demand and supply sides at national or international level. A much larger number of studies have provided situation analyses of human resources for health at national level in quite a wide range of countries. These latter studies provide descriptions relevant to an understanding of the whole health labor market, but often leave important gaps, such as information about salary and allowance levels or the extent of vacant posts, which, filled, would better allow for more comprehensive labor market analysis.

For example, Clark, Stewart, and Clark (2006) and Pond and McPake (2006) used international labor market frameworks to analyze the issue of health professional migration. Clark, Stewart, and Clark (2006) identified growing shortages, particularly of registered nurses in the United States and Europe, as the underlying force driving increased migration characterizing the decade up to 2005. While identifying factors such as aging and the growth of opportunities for women in alternative careers in both regions as underpinning the shortage, they failed to clarify why markets failed to clear, the explanation of which is likely to depend on the variable institutional arrangements for nursing (and other health professional) employment in different countries.

Under the heading of theories of globalization, Clark, Stewart, and Clark (2006) explored how health professional migration might represent the efficient operation of the global market in sourcing health labor formation inputs at least cost and in securing health professional labor at lower cost, and might in turn enable the labor forces of high income countries to compete in other markets. This assumes that market mechanisms operate and hence health care employers save on wages or other employment costs by expanding sources of labor supply to new geographic areas. The authors did not, however, provide data to support this theoretical argument.

Pond and McPake (2006) explored the political decisions that have shaped the demand side in the health labor markets of Organisation for Economic Co-operation and Development (OECD) countries, showing that political processes have tended to be cyclical in expanding and contracting demand for health professionals' labor. In the early to mid-1990s, expansion tendencies coincided across several large OECD countries (France, Germany, the United Kingdom, the United States), while in the preceding periods there had either been a tendency to constrain the growth of government-funded health professional training institutions in these countries or a tendency to diminish relative pay of some health professionals, resulting in reduced recruitment to training institutions. Importing health professionals was proposed as a short-term mechanism to manage resulting market disequilibrium.

Among national studies that used a health labor market framework explicitly, Vujicic *et al.* (2011) considered the market for physicians in Vietnam. They find

that the higher returns to urban public health sector employment compared with rural public health sector employment are influenced by the potential profitability of dual practice and the bonuses that can be earned from user-fee revenue sharing, which are most pronounced in higher level hospitals. These factors negate the impact of higher basic salary levels in rural areas. Their other main finding is a very low level of labor mobility, suggesting limited force of the market on employment decisions of Vietnamese doctors and contrasting with these doctors' stated employment objectives (to attain more urban, private, curative, and specialized posts) in an earlier phase of the same research project (Witter *et al.* 2011). The reasons why these doctors were not more mobile in pursuing their employment objectives are unclear.

Herbst *et al.* (2011) reviewed components of the health labor market for all types of health workers in Zambia using data largely related to 2005–08. The low stock of health workers in Zambia was explained by capacity constraints on health training institutions and out-migration that has been estimated well above the Sub-Saharan African average. Funded but unfilled posts suggest labor shortage market disequilibrium in the terms of Fields and Andalon (2008), although there were also further “established” posts that were not funded. Acute maldistribution of the health workforce was explained by unattractive market conditions in rural areas despite the introduction of a rural incentive scheme, and to a significant degree, unattractive nonfinancial market conditions were estimated to contribute to these.

Herbst *et al.* (2011) linked high absenteeism to the nonmarket forces of inadequate management and accountability mechanisms, but also the market related forces involved in the competing incentives of dual practice. They identified problems of quality and quantity of labor supply, but did not undertake further analysis of the roles played by capacity and motivational issues. They identified limited fiscal space and national health expenditure as factors contributing to problems across the health labor market both directly (in the declining wage bill over time, which was attributed largely to shifts in donor funding from general budgetary support) and indirectly (in affecting conditions of employment, particularly in rural areas that gave rise to unattractive nonfinancial market conditions there).

Appiah-Denkyira *et al.* (2013) reviewed components of the health labor market for all types of health workers in Ghana, relative to Zambia. They find that Ghana's health workforce is more plentiful, less badly distributed between urban and rural areas, and performing better on absenteeism and quality of care, although problems in all these areas exist. Informal charging seems to be limited. They argue that these outcomes stem from both labor market dynamics and the processes of the health professional education system, although they do not extensively identify the roles played by labor market forces.

Feyisa *et al.* (2012) provide a similarly structured review of the Ethiopian health labor market, explicitly identifying few specific market forces as influential. However, among those are the mechanisms operating in dual practice and



the greater potential earning level of private practice in urban areas, competition for health professionals from the private sector (with particularly large salary differentials for doctors), and motivational responses to perceptions of low salaries and poor working conditions.

### **Causal Labor Market Analysis**

Descriptive analysis alone cannot be used to inform specific policy design or examine the causal impact of policy change. Causal labor market analysis attempts to identify the causal effects of changes in labor market supply and demand conditions or regulation on the behavior of both employers and workers, and on the value of these changes to society. Understanding the range of factors that can influence labor market behavior is essential in developing policies to reduce labor market disequilibrium to improve societal outcomes. Behavior can be defined as labor market behavior (recruitment, retention, retirement, career choices, hours worked), labor productivity and health care outputs (e.g. number of services provided), or the value of changes in productivity to society (captured through wages where markets work, or in health care, measures of health status and quality of care). Theoretical models can help one understand the relationship between policy changes and behavior.

These types of empirical analyses may be conducted within a number of contexts from a full cost-benefit analysis of the costs and welfare effects of a specific labor market policy intervention, to the estimation of effects of such policies on a single labor market outcome, such as wages, retention, or distribution. The latter is more common in health and labor economics. Both are referred to as “program evaluation” and can be conducted retrospectively (after the policy is introduced and data collected) or prospectively (the policy has not yet been introduced and primary data collection may be required).

In health care, the most common types of cost-benefit analysis are concerned with skill mix changes that are evaluated using randomized controlled trials (Laurant *et al.* 2004). Typically these are not cost-benefit analyses as strictly defined as they measure benefits in nonmonetary terms using cost-utility or cost-effectiveness analysis, given the absence of prices that reflect consumers’ value of services provided in health care. These outcomes therefore include measures of quality of care and health outcomes of patients.

It is essential that information on both costs and benefits are used to decide on the value of different labor market policies. In health care, apart from the literature on skill mix change, there are very few studies that examine costs as well as benefits.

### ***How Is Causality Established?***

The central issue in causal labor market analysis is being able to move from examining an *association* between a policy intervention and an outcome—typically not useful for policy design—to measuring a *causal effect* such that we know with more certainty that changing *X* will lead to a change in *Y*.

Ideally, we would like to observe changes in the outcome ( $Y$ ) in the presence and absence of the labor market policy, then the difference in outcome between these two states of the world would reflect the effect of the labor market policy. This requires data on subjects  $i$  (individuals, organizations, or geographic areas) exposed to the policy intervention  $T_i$  (the treatment group) and subjects who are not  $C_i$  (the control group).

A causal effect is identified when the subjects in the treatment and control groups are identical with respect to all factors that might influence  $Y_i$ . The effect of the policy on  $Y$  is “unbiased.” All factors that might influence the outcome are controlled for and so are identical, and the only difference between the treatment and control groups is the presence of the labor market policy. If not, there is the possibility that a key factor that is different between the treatment and control groups has influenced the outcome, rather than labor market policy. The problem is that only some of the factors that influence  $Y_i$  are observed in the data ( $X$ )—other factors influencing  $Y_i$  ( $Z$ ) might be unobserved to the researcher. Knowledge of the process through which subjects are assigned to the treatment and control groups is therefore crucial in identifying a causal effect. Were subjects assigned by chance (randomization) or by some other selection process that was or can be observed?

Identifying causal effects is often sought through different types of study design and the use of micro-econometric analysis. Different types of analysis use different methods to try and identify the counterfactual, i.e. what would have happened to  $Y$  in the absence of the labor market policy. The outcome may change for other reasons that need to be accounted for, which is often referred to as endogeneity, where the effect of the policy on  $Y$  is partly determined by other unobserved factors that influence both the policy and the outcome  $Y$ . In this case, we cannot be sure whether the policy or the unobserved factor influenced  $Y$ .

### ***Experimental Approaches***

Experimental study designs such as randomized controlled trials, where randomization helps ensure that subjects are identical in the treatment and control groups, are regarded as the “gold standard” of study design, and are a relatively recent development in labor economics, where they are referred to as field experiments (List and Rasul 2011). Though use of randomized trials in health care is much more common, their application in evaluating labor market policies is limited to skill mix change, partly because randomization may not be possible where labor market policy is changed for a whole country—thus a concurrent control group cannot be used and a counterfactual cannot be identified.

Quasi-experimental designs include difference-in-difference analysis where observations on subjects ( $i$ ) in the treatment ( $T$ ) and control ( $C$ ) groups are made before ( $Y_{t-1}$ ) and after ( $Y_{t+1}$ ) the policy is introduced, often for a number of periods so time trends can be captured (Angrist and Krueger 1999). One can then observe the difference in the outcome for the treatment group over time,

$Y_{t+1}^T - Y_{t-1}^T$ , and the difference in the outcome for the control group (the counterfactual) over time,  $Y_{t+1}^C - Y_{t-1}^C$ . Under certain assumptions, the difference between these differences— $(Y_{t+1}^T - Y_{t-1}^T) - (Y_{t+1}^C - Y_{t-1}^C)$ —is the effect of the policy intervention on  $Y$ . This is implemented within an econometric framework that depends on the nature of the data at hand. Any differences in the observed characteristics of the control and treatment groups can be included as independent variables in the regression model.

Although randomization should be used wherever possible, there are often good reasons to use other study designs where randomized trials are not possible, do not replicate real-world implementation of a policy, are subject to bias in their design and selection of subjects, or are not generalizable to the population (Rothwell 2005).

Most empirical approaches in labor economics use revealed preference data on actual behavior change. Labor economists often mistrust data on stated intentions. However, in health economics and other areas where market failure is pervasive (e.g. environmental economics) and there are simply no data on revealed preferences, stated preference methods are increasingly used in health workforce surveys. In particular, discrete choice experiments (DCEs) are based in random utility theory and can provide important data on preferences on which to base policy. An advantage of stated preference methods is that choices are presented exogenously to respondents using the experimental design, and so the difficulties of identifying causal effects of job characteristics are largely avoided. We therefore classify DCEs in the general class of experimental methods, but they are more similar to laboratory experiments than field experiments.

### ***Non-Experimental Approaches***

More often than not, experimental designs are infeasible and so micro-econometric methods are used to attempt to control for unobservable factors that influence  $Y$ , in addition to the labor market policy. The aim of using a control group that has not been exposed to the intervention (the counterfactual) is still the key feature of the study design. The problem in non-experimental methods is that the control group is not identical to the treatment group in all respects. Though one might observe a difference in  $Y$  if a policy changes, this may in fact be due to other factors that determine  $Y$  and are correlated with  $X$ . For example, in a labor supply model that attempts to examine the effect of a change in wages on hours worked ( $Y$ ), another variable that influences both wages and hours worked is gender. Women earn less than men and work fewer hours. The size of the association between wages and hours (i.e. the wage elasticity) also includes, and is confounded by, the effect of gender on hours worked. The wage elasticity is therefore biased. Factors that influence  $Y$  apart from labor market policy and are observed in the data ( $X$ ) can simply be included as independent variables in the regression model, reducing the bias of the effect of the policy change. Controlling for

confounding variables is an essential element of regression analysis (Angrist and Krueger 1999).

The next source of bias is that there may also be factors ( $Z$ ) that influence  $Y$  and  $X$  but are not observed in the data. This is an important problem in micro-econometrics that has been handled in two main ways. The mechanism that assigns a subject to the labor market program is not observed or measured.

The first way is where  $Z$  does not change over time—it is time invariant. In the labor supply model, one might assume that unobserved ability and skills of workers or the characteristics of firms they work at influence both wages and hours worked but are fixed over at least short periods. The use of panel data where  $Y$ ,  $X$ , and  $Z$  of workers are observed over multiple time points provides an important opportunity to control for variables that are time invariant. Each observation is no longer independent of the next, a key assumption of linear regression analysis. A single individual will have multiple observations of  $Y$  and  $X$  in the data, one observation for each period.

Panel data econometrics techniques account for  $Z$  using two main methods. The first method (random-effects panel data models) accounts for the correlations of  $Z$  and other unobserved factors influencing  $Y$  within the same individual over time or of individuals within the same firm. The second method (fixed-effects panel data models) exploits the fact that if one were to take first differences of  $Y$  and  $X$  (e.g.  $Y_{it} - Y_{it-1}$ ) and so use changes over time as variables in the regression model, then the difference in any value of  $Z$  that does not change over time is zero, and so does not influence the change in  $Y$ . The  $X$  that we observe (e.g. gender) and the  $Z$  that we do not observe (e.g. ability and skills) do not need to be included in the model. Fixed-effects methods also have the advantage of allowing for correlations between the unobservable fixed effects and all other independent variables in the regression model.

Panel data econometrics and its many variants are key tools in micro-econometrics to estimate close to causal effects. Panel data techniques can also be used alongside difference-in-difference study designs that rely on panel data with multiple measurements at different points in time.

The collection of panel/longitudinal data is very important. Not only can it be used to apply panel data econometrics, but panel data can be used to account for the timing of the introduction of a labor market policy to avoid a common problem with cross-sectional data collected at a single point in time: reverse causality. In a labor supply model, wages influence hours worked, but hours worked also influence wages. In a cross-section of data, it is not possible to determine the direction of causality. However, this is essential to estimate causal effects where the labor market policy is introduced first (e.g. an increase in wages), which consequently influences behavior (increased hours worked). With panel data one can account for reverse causality by using the values of variables measured at different points in time, such that wage change is

measured now when it occurs, and hours worked are measured at future time points. Using the lagged value of wages rather than the current value can account for reverse causality.

The second main way to reduce bias caused by unobserved  $Z$  is instrumental variable (IV) estimation. This technique explicitly recognizes that the presence of the labor market policy is influenced by a specific set of factors that are different from the factors that influence the outcome  $Y$ . This may be when the labor market policy is introduced but where take-up of the policy is voluntary and subject to some type of selection process. A pay-for-performance program may be voluntary, union membership is voluntary, as might a move to a rural area that now attracts financial incentives. Even where a hospital introduces such a program for all its workers, workers can choose to leave or join this hospital after the policy is introduced, meaning that the measured effect of the program on performance may not be due to the program itself, but a different mix of workers at the firm or some other factors that influence the choice of hospital.

In the case of IV, we not only know that  $Z$  is correlated with  $Y$  and  $X$ , but we wish to model this relationship explicitly, often in a separate regression model. Furthermore, we need to choose an independent variable (the IV) that influences the adoption of the policy (e.g. that influences workers choosing to participate) and so causes variation in the adoption of the policy, but that is not correlated with unobserved factors that influence the outcome  $Y$ . The IV needs to be exogenous to  $Y$  and not be correlated with it. Ideally, one needs to observe something else that influences the decisions of health workers to join the program or make use of the labor market policy that has nothing to do with the labor market policy or its outcomes. This therefore attempts to identify a causal effect of the policy on  $Y$ , by separating out the part of the variation in  $Y$  that is due to the effect of the exogenous variation driven by the IV. Finding a variable that influences selection into the program but that does not influence the outcome of the program is often problematic, and a large literature discusses how this can be done (Angrist and Krueger 1999).

Other types of analysis include regression discontinuity designs and matching methods. Regression discontinuity designs exploit abrupt changes in  $Y$  over time, which control for otherwise smooth trends. The allocation of subjects to the treatment or control group is determined by a threshold and the movement of subjects across the threshold (DiNardo and Lee 2011). This may include eligibility for a specific payment being determined by a specific level of measured performance or other measurable criteria. Identification of causal effects comes from the fact that subjects just below the threshold will be very similar to those just above the threshold—and that their position just above or below the threshold is effectively random. This relies on good knowledge of the process used to determine the threshold.

Matching methods is another set of methods used to construct a control group that closely matches the characteristics of the treatment group. It provides a

more comparable control group than regression methods, though it relies only on the characteristics of the subjects that can be observed.

Many of the above methods can also be combined in a single analysis where the data allow. For example, fixed effects, reverse causality, and IVs can be combined in a dynamic panel data model. This also recognizes that  $Y$  (e.g. hours worked) this year is influenced by  $Y$  in previous years and so also examines persistence while using fixed effects (first differences) and lags of independent variables as instruments (Bond 2002).

## Data Requirements

Ideally, data are collected that are relevant to the research question being addressed. This approach is possible with specially designed surveys and data collection exercises, but where it is not, one needs to rely on existing administrative and survey data. A variety of sources contain data on the health workforce, including personnel records of employers, registration and accreditation databases, tax records, or health insurance records. Labor force surveys usually contain richer information on worker characteristics than administrative sources.

The previous section highlighted the importance for causal labor market analysis of panel data, where measurements on individuals are linked over time in panel or cohort studies. With only a cross-section, it is much harder to identify causal effects.

As said, data on wages and earnings of health workers are essential for any kind of economic labor market analysis. Changes and movements in wages and earnings, combined with other information, tell us much about how market forces are operating and the behavior of workers. The centrality of their role in labor markets, even in those that do not work very well, means that most types of analysis in labor economics involve the collection and analysis of data on wages or earnings of health professionals, and examining their differentials between occupations, specialties, age groups, genders, and geographic areas (including countries) reveals much about how labor markets are working. Unfortunately, many health care systems have failed to translate this fundamental requirement for labor market analysis into data collection items in surveys of the health workforce or in administrative data collections.

Descriptive data are important in describing what is currently happening, or what has happened in the past. As a minimum, these datasets often include the number of health workers of different types: age, gender, hours worked, specialty, and geographic location. These data are usually obtained from registration databases or personnel records. If data come from personnel records of health systems, they may also include information on wages and earnings, which are essential for any type of basic labor market analysis. Cross-tabulations of data (e.g. wages by specialty) can be used to begin examining associations between key variables, which can in turn be used to generate and

test hypotheses as to why changes have occurred, or why the current labor market is as it is.

Data that are used to examine the causal effects of policy change include the same descriptive data, but often differ in two aspects. First, they contain more variables to enable the researcher to control for as many factors influencing behavior as possible. These include data linked to health worker organizations and their characteristics, and information about their household (household income and the presence of children are predicted in labor models to have a large influence on hours worked), which are necessary to isolate the causal effect of policy, rather than just showing association. These data might include information from surveys and primary data collection that obtain information about those family circumstances that are not usually available in administrative data sources.

Second, they can be longitudinal, with health workers followed up over time and appearing in the data every year. This requires a unique identifier in datasets that can be used to link individuals over time. Such panel data are essential in being able to control for unobserved factors that do not change over time, and in accounting for reverse causality.

Administrative datasets are usually comprehensive in coverage of the population of health workers, and may have data on health workers organized in firms (e.g. hospitals and their characteristics) but often contain little data on health worker characteristics or their productivity or performance. Data on health workers may be linkable over time.

Surveys of the health workforce can enable the collection of much richer data on the characteristics of health workers, their jobs, their families, and the characteristics of where they live. One can collect “bespoke” data that directly address the research question being asked. However, a key issue with surveys is their external validity and generalizability, so it is important to use appropriate sampling methods and survey weights to assess representativeness.

The linkage of health workforce data (inputs) to data on outputs and outcomes in health care (see figure 2.1) is difficult in many countries. Yet such linkage is fundamental for examining the impact of different health workforce policies on efficiency and equity in health care. These datasets go beyond health workforce datasets. For example, a dataset of hospital activity and outputs should ideally link each episode of care to the health worker who provided that episode of care, and his or her characteristics. The design of administrative datasets that, within a hospital, link personnel records to hospital activity and outputs would provide a key link between health workers and the outcomes they produce, greatly facilitating research and analysis of the most efficient use of the health workforce.

An example of a dataset designed specifically to analyze medical labor markets using labor economics approaches is the Medicine in Australia: Balancing Employment and Life (MABEL) panel survey of doctors (box 4.1).



### Box 4.1 MABEL Panel Survey of Doctors

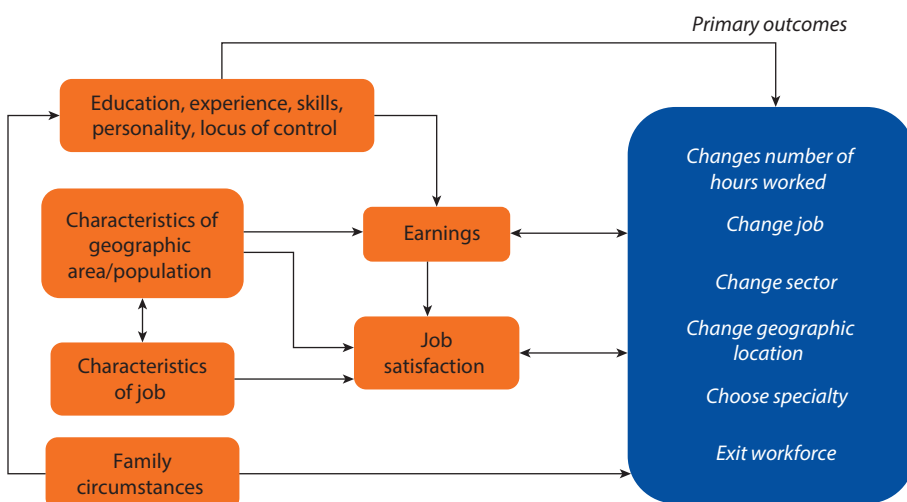
The MABEL survey is an example of a panel survey that collects data from around 10,000 doctors (20% of all doctors in Australia) every year.<sup>a</sup> It is unique as it is the only panel survey in the world examining physicians that has been designed for examining causal effects through the use of panel data econometric analysis. The survey can also be used for descriptive analysis.

The detailed methods and design of the survey including copies of questionnaires are available on the website ([www.mabel.org.au](http://www.mabel.org.au)). The survey is very much focused on the supply side as it is a survey of the decisions and characteristics of doctors, and so contains little information about need or demand. The survey included DCEs in Wave 1 (2008) and can be used to observe doctors' actual job changes over time. MABEL is funded by the National Health and Medical Research Council, through the Centre for Research Excellence in Medical Workforce Dynamics. The survey is currently funded for nine annual waves through 2016.

There are three research themes that MABEL is used to address: workforce participation (e.g. labor supply models), career transitions (e.g. specialty choice and retirement, retention), and rural workforce supply and distribution. The research also looks at the role of financial incentives and other factors influencing labor market decisions and outcomes. In addition to collecting data on earnings, the survey collects a range of information on family circumstances including information about the doctor's spouse and children, which can have important effects on hours worked and on choosing whether to work in a rural area (Figure B4.1.1 provides an overview of the type of data collected).

The dataset relies on a national source of data containing the contact details of all doctors in Australia that can be linked over time. This also enables the calculation of response weights that help ensure that the data are representative of the population of doctors in Australia. Research findings from MABEL can be found on the website.

**Figure B4.1.1 Types of Data Collected by the MABEL Survey**



a. There are, of course, other panel datasets on the health workforce in other countries, but these are administrative datasets rather than survey datasets, with few including information on earnings.



## References

- Angrist, J. D., and A. B. Krueger. 1999. "Empirical Strategies in Labor Economics." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 1277–366. Elsevier.
- Appiah-Denkyira, E., C. H. Herbst, A. Soucat, C. Lemiere, and K. Saleh. 2013. *Towards Interventions on Human Resources for Health in Ghana: Evidence for Health Workforce Planning and Results*. Washington, DC: World Bank.
- Blundell, R., and T. Macurdy. 1999. "Labor Supply: A Review of Alternative Approaches." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 1559–695. Elsevier.
- Bond, S. 2002. "Dynamic Panel Data Models: A Guide to Micro Data Methods and Practice." *Portuguese Economic Journal* 1: 141–62.
- Clark, P. F., J. B. Stewart, and D. A. Clark. 2006. "The Globalization of the Labour Market for Health-care Professionals." *International Labour Review* 145.
- DiNardo, J., and D. S. Lee. 2011. "Program Evaluation and Research Designs." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 463–536. Elsevier.
- Feysia, B., C. H. Herbst, W. Lemma, and A. Soucat. 2012. *The Health Workforce in Ethiopia: Addressing the Remaining Challenges*. Washington, DC: World Bank.
- Fields, G., and M. Andalon. 2008. "A Toolkit for Analyzing Labor Markets in the Health Care Sector in Africa." Health, Nutrition, and Population, AFTHD Africa Region. World Bank, Washington, DC.
- González, P., and I. Macho-Stadler. 2013. "A Theoretical Approach to Dual Practice Regulations in the Health Sector." *Journal of Health Economics* 32: 66–87.
- Herbst, C. H., M. Vledder, K. Campbell, M. Sjöblom, and A. Soucat. 2011. *The Human Resources for Health Crisis in Zambia: An Outcome of Health Worker Entry, Exit, and Performance within the National Labor Health Market*. Africa Human Development Series. Washington, DC: World Bank.
- Laurant, M., D. Reeves, R. Hermens, J. Braspenning, R. Grol, and B. Sibbald. 2004. "Substitution of Doctors by Nurses in Primary Care." *Cochrane Database of Systematic Reviews* 4.
- List, J. A., and I. Rasul. 2011. "Field Experiments in Labor Economics." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 103–228. Elsevier.
- Moffitt, R. A. 1999. "New Developments in Econometric Methods for Labor Market Analysis." In *Handbook of Labor Economics*, edited by O. C. Ashenfelter and D. Card, 1367–397. Elsevier.
- Pond, B., and B. McPake. 2006. "The Health Migration Crisis: The Role of Four Organisation for Economic Co-operation and Development Countries." *The Lancet* 367: 1448–55.
- Rothwell, P. M. 2005. "External Validity of Randomised Controlled Trials: 'To whom Do the Results of this Trial Apply?'" *The Lancet* 365: 82–93.
- Scott, A., P. Sivey, C. Joyce, D. Schofield, and P. Davies. 2011. *Alternative Approaches to Health Workforce Planning*. Report by the National Health Workforce Planning and Research Collaboration, Health Workforce Australia, Adelaide.

- Vujicic, M., B. Shengelia, M. Alfano, and H. B. Thu. 2011. "Physician Shortages in Rural Vietnam: Using a Labor Market Approach to Inform Policy." *Social Science & Medicine* 73: 970–77.
- Witter, S., B. Thi Thu Ha, B. Shengalia, and M. Vujicic. 2011. "Understanding the 'Four Directions of Travel': Qualitative Research into the Factors Affecting Recruitment and Retention of Doctors in Rural Vietnam." *Human Resources for Health* 9: 20.

# Gaps and Future Research

*Of 57 countries in Africa, only around one-third of them have featured in this review, suggesting major geographic gaps in research on health labor markets in crisis-affected countries. The main topical gap is the limited link between theory and empirical evidence. Key aspects of the health labor market, such as labor demand and its determination, informal charges, and others remain unexplored. The literature using discrete choice experiments to understand health workers' preferences for job attributes has grown considerably, but stated choices need to be validated to allow generalization of results.*

The academic literature that studies the health workforce using economic methods of analysis is small. The most numerous types of studies are probably studies of the labor supply of nurses that attempt to estimate wage elasticities (Antonazzo *et al.* 2003; Shields 2004). There is also a relatively large literature evaluating pay for performance. Most of these studies are from the United States or Europe. There is also a growing literature that uses discrete choice experiments (DCEs) to examine the job choices of health workers, most of it for low-income countries.

## Gaps by Topic

An absence of material on labor demand and how it is constituted and determined is a marked topical gap in the review. Vujicic, Addai, and Bosomprah (2009) provide a good starting point, reviewing how the public sector wage bill is determined. Public sector wage bills combine with other sources of effective demand including elements of the Ministry of Health's funds that support health worker pay and employment (such as allowances and performance related bonuses) and private sector demand and expenditure on public and private services. These are needed before the demand side of health labor markets can be fully described, much less analyzed. The distribution of demand appears to be a major factor in shaping labor supply responses, and it is only in understanding the distribution across both public and private sectors that this can be clearly

seen, as evidenced by a range of studies that show that the opportunities for dual practice in urban areas shapes the supply response despite public sector remuneration favoring rural areas in some cases.

Qin, Li, and Hsieh (2013) was the only study to estimate a labor supply model. They demonstrate the difficulties of estimation of the elasticity of labor supply without reliable and large-scale data sets but suggest that census data in countries in which particular questions are included (concerning field of employment, professional qualifications, salary/wage, and number of hours worked) might enable similar analyses in other countries where there is significant salary or wage rate variation across employees in different sectors or regions.

The one cross-country based analysis of the elasticity of retention (or the responsiveness of out-migration to pay differences between countries) failed to find a strong relationship (Vujicic *et al.* 2004). This might be because health worker categories were not sufficiently disaggregated. It could also be because wage differentials were measured in absolute terms, whereas it could be that health workers are most responsive to relative rates of remuneration and may have reservation wages defined by local social status associations. Alternatively, it could be because it is necessary to measure other “push” factors as they apply differentially across countries suffering out-migration and found by a number of studies to be equally important to remuneration level (for example, Mangham and Hanson 2008; Penn-Kekana *et al.* 2005). These alternative explanations provide a potential agenda for future research.

There is now a quite considerable body of studies using DCEs to understand health professionals’ valuations of different job attributes and these appear to provide valuable information for policy makers. But the variability of responses suggests that generalizable conclusions will require more research, although the studies that have analyzed the correlates of utility measures are a good start.

Given the variability, there is value in adding studies by country and cadre for local policy to be formulated, but these studies are complex and there is limited capacity at country level to undertake such studies appropriately. Before undertaking a wholesale and expensive effort to repeat DCEs for further cadres and geographic settings, further research should be considered, first to establish whether the responses to hypothetical questions can reliably predict behavior, and second to explore whether there is a simpler way of establishing the most important job attributes in different settings. Establishing correlations between DCE results and simpler survey results could enable a wider range of studies to be undertaken using simpler surveys, at lower cost.

The dual practice literature was critiqued as providing limited linkage between theory and empirical evidence, with very little of the latter. There is therefore considerable scope for empirical work that will test existing theories and generate new propositions about the relationships between key variables involved in dual practice such as the relative profitability of public and private sector work, the

dependency of private income streams on public sector job holding, and the undertaking of private practice within and outside public service facilities.

The significance of the findings of highly variable health worker effort associated with institutional characteristics (even if variably so) is that they provide support for the operation of a significant role of the health labor market in shaping motivation, effort, and performance. If, as appears to be the case in Delhi and in Arusha (Das and Hammer 2004, 2005; Leonard and Masatu 2010), nonpublic facilities or facilities that have greater control over price and quantity (hiring and firing, salary setting) better elicit greater effort, this might be interpreted as having implications for the role that stronger incentives can play in determining quality of labor supply as well as quantity. Evidence in Das and Sohnesen (2007) tempers the conclusion that might otherwise emerge that the private sector is always better able to manage market forces toward greater effort, suggesting perhaps that as health systems develop, public institutions can better manage outcomes.

The issue is not about whether there are public or private providers, but the incentives that each contains for effort and productivity and efficiency. Further research focusing on the processes shaping effort levels could clarify not only how this works in the private sector for richer patients in contexts such as Delhi, but how it can work to protect the interests of poorer patients who will not be able to access such facilities in the public sector in contexts such as Paraguay.

Data on salaries and allowances in the public sector and on remuneration in general in the private sector were demonstrated by McCoy *et al.* (2008) to be incomplete and barely capable of enabling even simple analyses of trends in absolute and relative levels, and in pay structures. As pay is the critical variable in any labor market analysis, advances in the understanding of labor markets in low- and middle-income countries (LMICs) will not be possible without significant improvements in data availability on the critical price variable.

Given the complexity introduced by the important contribution made by allowances, and the distribution of these among different health workers, a data collection system that can provide an adequate understanding of public sector pay will also need to be quite complex. Tracking private sector salaries and self-employed earnings and earning potential is even more challenging, but examples of large-scale earnings and salary tracking exercises are available in other industries<sup>1</sup> and in this, health may not be significantly more complex than other sectors. It would probably not be feasible, however, to identify the full return to public health sector employment in an internationally comparable way, taking account of supplementary earnings that are derived or related to employment in a public role.

Further exploration of the existence of informal charges in private institutions as well as public institutions in a wider range of geographic settings would enhance understanding of the return to private sector employment, and whether it too is underestimated by salary data. Further work on informal charges might clarify the labor market conditions that promote and restrict

informal charging, especially if Barr, Lindelow, and Serneels's (2009) conclusion that wage rates do not provide strong explanations can be verified in data from real labor markets.

On the basis of the argument of Pond and McPake (2006), health professional migration would be predicted to rise and fall with the political cycles of larger Organisation for Economic Co-operation and Development (OECD) countries, whereas the argument of Clark, Stewart, and Clark (2006) predicts the migration trend would only abate as training institution cost differentials and salary differentials between rich and poor countries diminished. This provides a potential agenda for research clarifying the relationships between overall migration levels, patterns among source and destination countries, and labor formation costs and local market conditions in both source and destination countries.

### **Geographic Gaps**

Of 57 countries in Africa—36 deemed to be in human resource crisis (WHO 2006)—only 17 feature in the References, with sole pieces of evidence or studies available for 7 of these 17 and only 2 pieces of evidence or studies for a further 6. This suggests major geographic gaps in research on health labor markets in crisis-affected countries and suggests a need for a greatly extended research effort (also, see figure 1.1).

It is also of note that among the countries in which no or very limited research has been identified are the most fragile and conflict-affected states. No studies were found, for example, in Angola, Sierra Leone, the Democratic Republic of Congo, Sudan, or Somalia, and only one in Liberia—perhaps the most conflict-affected set of countries in the continent. If human resource scenarios are distinct in such settings, this presents a significant gap in our understanding of labor markets.

### **The Way Forward**

This document has provided an overview of the application of economics to health worker labor markets. This framework needs to be applied much more consistently and systematically to understand how health care labor markets work, the extent and impact of market failure and market imperfections, and the types of analysis and data required to undertake economic analysis of health human resources.

One potential reason why economics has not been applied is the lack of data, especially on earnings. Improved datasets are an essential starting point that will help to urgently move beyond counting the health workforce and some of its basic characteristics, to understanding the determinants and solutions to labor market disequilibrium using high-quality descriptive and causal evidence. Administrative and survey datasets need to be designed so they can be matched to each other, matched over time, and also matched to data on health care utilization and health system performance and outcomes. A final way forward is to

help train and attract more labor and health economists to undertake this vital research. Building capacity in health workforce research is fundamental in taking this agenda forward.

## Note

1. For example, the International Labour Organization's Key Indicators of the Labour Market database contains average monthly wages in the total economy for 99 countries, and average compensation costs for the employment of workers in the manufacturing sector for 34 countries. <http://kilm.ilo.org/2011/download/kilm15EN.pdf> (accessed January 1, 2013).

## References

- Antonazzo, E., A. Scott, D. Skatun, and R. F. Elliott. 2003. "The Labour Market for Nursing: A Review of the Labour Supply Literature." *Health Economics* 12: 465–78.
- Barr, A., M. Lindelow, and P. Serneels. 2009. "Corruption in Public Service Delivery: An Experimental Analysis." *Journal of Economic Behavior and Organization* 72: 225–39.
- Clark, P. F., J. B. Stewart, and D. A. Clark. 2006. "The Globalization of the Labour Market for Health-care Professionals." *International Labour Review* 145.
- Das, J., and J. Hammer. 2004. "Which Doctor? Combining Vignettes and Item-Response to Measure Doctor Quality." Policy Research Working Paper 3301, World Bank, Washington, DC.
- . 2005. "Money for Nothing: The Dire Straits of Medical Practice in Delhi, India." Policy Research Working Paper 3669, World Bank, Washington, DC.
- Das, J., and T. P. Sohnesen. 2007. "Variations in Doctor Effort: Evidence from Paraguay." *Health Affairs (Millwood)* 26: w324–37.
- Leonard, K. L., and M. C. Masatu. 2010. "Using the Hawthorne Effect to Examine the Gap between a Doctor's Best Possible Practice and Actual Performance." *Journal of Development Economics* 93: 226–34.
- Mangham, L. J., and K. Hanson. 2008. "Employment Preferences of Public Sector Nurses in Malawi: Results From a Discrete Choice Experiment." *Tropical Medicine and International Health* 13: 1433–41.
- McCoy, D., S. Bennett, S. Witter, B. Pond, B. Baker, J. Gow, J. G. S. Chand, T. Ensor, and B. McPake. 2008. "Salaries and Incomes of Health Workers in Sub-Saharan Africa." *The Lancet* 371: 675–81.
- Penn-Kekana, L., D. Blaauw, K. S. Tint, D. Monareng, and J. Chege. 2005. *Nursing Staff Dynamics and Implications for Maternal Health Provision in Public Health Facilities in the Context of HIV/AIDS*. Frontiers in Reproductive Health, Population Council.
- Pond, B., and B. McPake. 2006. "The Health Migration Crisis: The Role of Four Organisation for Economic Co-operation and Development Countries." *The Lancet* 367: 1448–55.
- Qin, X., L. Li, and C.-R. Hsieh. 2013. "Too Few Doctors or Too Low Wages? Labor Supply of Health Care Professionals in China." *China Economic Review* 24: 150–64.
- Shields, M. A. 2004. "Addressing Nurse Shortages: What Can Policy Makers Learn from the Econometric Evidence on Nurse Labour Supply?" *Economic Journal* 114 (499): F464–98.

- Vujcic, M., E. Addai, and S. Bosomprah. 2009. "Measuring Health Workforce Productivity: Application of a Simple Methodology in Ghana." Health, Nutrition and Population (HNP) Discussion Paper, World Bank, Washington, DC.
- Vujcic, M., P. Zurn, K. Diallo, O. Adams, and M. R. Dal Poz. 2004. "The Role of Wages in the Migration of Health Care Professionals from Developing Countries." *Human Resource for Health* 2: 3.
- WHO (World Health Organization). 2006. *The World Health Report: Working Together for Health*. Geneva.





## **Environmental Benefits Statement**

The World Bank Group is committed to reducing its environmental footprint. In support of this commitment, the Publishing and Knowledge Division leverages electronic publishing options and print-on-demand technology, which is located in regional hubs worldwide. Together, these initiatives enable print runs to be lowered and shipping distances decreased, resulting in reduced paper consumption, chemical use, greenhouse gas emissions, and waste.

The Publishing and Knowledge Division follows the recommended standards for paper use set by the Green Press Initiative. Whenever possible, books are printed on 50 percent to 100 percent postconsumer recycled paper, and at least 50 percent of the fiber in our book paper is either unbleached or bleached using Totally Chlorine Free (TCF), Processed Chlorine Free (PCF), or Enhanced Elemental Chlorine Free (EECF) processes.

More information about the Bank's environmental philosophy can be found at <http://crinfo.worldbank.org/wbcrinfo/node/4>.



Studying health labor markets more closely could help resolve globally widespread human resources for health problems. The lack of labor market analysis of these problems provides some explanation for their persistence. Policy makers in countries promulgating or refining strategies for achieving universal health coverage will find it important to understand how key elements in their health labor market are likely to interact and how these interactions could help—or hinder—progress toward universal health coverage. These interactions are complex and multidimensional, and *Analyzing Markets for Health Workers: Insights from Labor and Health Economics* highlights areas where forces in the health labor market matter most.

This book provides an overview of the key issues policy makers will encounter when attempting to apply economics to the analysis of health workers' labor markets. Though much has been written and planned about human resources for health, a major weakness with most of this analysis is that it does not use an economic perspective. The use of an explicit economic framework moves the focus away from simplistic but costly policy responses (such as training more doctors and nurses) toward understanding more carefully the role of incentives, productivity, and the distribution of health workforce. The health workforce is only one part of the health system, and a focus of analysis on only the health workforce will be insufficient to determine the optimal number of health workers. Due to well-recognized market failures in health care, market forces cannot be relied upon to solve well-known health workforce challenges, such as shortages, maldistribution, and performance.

Of interest to policy makers, researchers, and practitioners, *Analyzing Markets for Health Workers: Insights from Labor and Health Economics* is part of a multiyear World Bank Group program on human resources for health policies. The program's ultimate objective is to strengthen knowledge and capacity to collect evidence, analyze, and evaluate the effectiveness of human resources for health interventions in the context of a country's health system strengthening strategy and universal health coverage goals. It specifically addresses the theoretical and empirical evidence on health labor markets and proposes an agenda to address knowledge gaps and to build capacity for applying health labor market analysis in the context of low- and middle-income countries.



THE WORLD BANK

ISBN 978-1-4648-0224-9



SKU 210224